Lawrence Livermore National Laboratory

Advanced Simulation and Computing

FY14 Program Notebook for Computational Systems and Software Environment and Facility Operations and User Support

March 27, 2014

Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the U.S. Department of Energy, National Nuclear Security Administration under Contract DE-AC52-07NA27344.

LLNL-TR-652433

ASC™

Contents

Introduction

Mission and Vision

TBD

Requirements and Drivers

TBD

Program Structure

TBD

Five-Year Plans for Projects

TBD

Computational Systems and Software Environment (WBS 1.5.4)

The mission of this national sub-program is to build integrated, balanced, and scalable computational capabilities to meet the predictive simulation requirements of the National Nuclear Security Administration (NNSA). This sub-program strives to provide users of ASC computing resources a stable and seamless computing environment for all ASC-deployed platforms. Along with these powerful systems that ASC will maintain and continue to field, the supporting software infrastructure that CSSE is responsible for deploying on these platforms includes many critical components, from system software and tools, to input/output (I/O), storage and networking, to post-processing visualization and data analysis tools. Achieving this deployment objective requires sustained investment in applied R&D activities to create technologies that address ASC's unique mission-driven needs for scalability, parallelism, performance, and reliability.

Commodity Technology Systems (WBS 1.5.4.8)

This level 4 product provides production platforms and integrated planning for the overall system architecture commensurate with projected user workloads. The scope of this product includes strategic planning, research, development, procurement, hardware maintenance, testing, integration and deployment, and quality and reliability activities, as well as industrial and academic collaborations. Projects and technologies include strategic planning, performance modeling, benchmarking, procurement and integration coordination, and installation. This product also provides market research for future Commodity Technology Systems (CTS).

Production Planning and Integration (LLNL)

The LLNL ASC strategy for commodity technology systems is to leverage industry advances and open source software standards to build, field, and integrate Linux clusters of various sizes into classified and unclassified production service. The programmatic objective is to dramatically reduce overall total cost of ownership of these commodity systems relative to best practices in Linux cluster deployments today. This objective strives to quickly make these systems robust, useful production clusters under the coming load of ASC scientific simulation capacity workloads.

Required Capabilities

R1: Robust commodity Linux compute clusters to run ASC simulations

Gaps

G1: Rapidly evolving technology in the commodity space necessitates agility in all facets of cluster deployments (hardware, software, facilities)

Five-Year Plan

FY14

- Provide production support for TLCC2 systems
- Lead the tri-lab process to conduct a commodity technology system (CTS)-1 market survey and develop the RFP
- ~~Monitor computer industry developments for opportunities to enhance capacity computing and the associated infrastructure (G1)~~
- ~~Plan for procurement of TLCC3 systems~~

FY15

- Monitor computer industry developments for opportunities to enhance capacity computing and the associated infrastructure (G1)
- Plan for deployment of CTS-1 ~~TLCC3~~ systems

FY16

- Monitor computer industry developments for opportunities to enhance capacity computing and the associated infrastructure (G1)
- Deploy CTS-1 ~~TLCC3~~ systems

FY17

- Monitor computer industry developments for opportunities to enhance capacity computing and the associated infrastructure (G1)

FY18

- Monitor computer industry developments for opportunities to enhance capacity computing and the associated infrastructure (G1)
- Plan for procurement of CTS-2 systems

Computing Platform Integration and Deployment (LANL)

The scope of the Computing Platform Integration and Deployment project is to accept delivery and begin deployment of production systems. Primary capabilities include completing the acceptance tests, diagnostics tests, integrating the systems into the LANL yellow network, system stabilization, and transition into the classified network.

Required Capabilities

R1: Understanding of industry roadmaps and technologies available in the CTS-1 ~~TLCC3~~ timeframe

R2: Working with vendors and users, identify potential architectures to meet mission need

Gaps

G1: Technology roadmaps line up with ASC CTS ~~capacity~~ platform requirements

Five-Year Plan

FY14

- Provide production support for TLCC2 systems
- Participate in the tri-lab planning for the NNSA ASC CTS-1 ~~TLCC3 review and platform selection (G1)~~
- Continue to operate Luna and the other capacity systems in both the classified and unclassified computing environments

FY15

- Provide production support for TLCC2 systems
- Participate in the tri-lab planning for the NNSA ASC CTS-1 (G1)
- Integrate CTS-1 ~~TLCC3~~ systems into LANL red and turquoise networks

FY16

- Provide production support for CTS-1 ~~TLCC3~~ systems
- Decommission TLCC2 systems

FY17

- Provide production support for CTS-1 ~~TLCC3~~ systems

FY18

- Provide production support for CTS-1 system (G1)

ASC Commodity ~~Capacity~~ Systems (SNL)

The purpose of the ASC Commodity ~~Capacity~~ Systems project is to support the acquisition, delivery, and installation of new ASC commodity technology ~~Capacity~~ systems. The project is supported by analysis of SNL's portfolio of application needs for capacity workload ~~computing~~ systems within the context of the higher integrated ASC platform strategy of commodity and advanced technology systems. Efforts include definition of requirements for TLCC systems and collaboration with the CCE product, with respect to a common software stack for new and existing capacity systems.

Required Capabilities

R1: Provide production computing cycles for ASC applications, especially those that for technical or programmatic reasons are not expected to run on advanced technology systems

Gaps

G1: The demand for computing cycles continues to increase beyond whatever is available

G2: Future CTS will likely include advanced architecture computing capabilities, for example GP-GPU accelerators; there is a significant gap in the ability of ASC applications to use this performance potential and a related gap in memory performance/capacity

Five-Year Plan

FY14

- Monitor industry hardware and software roadmaps to understand opportunities for cost-effective integration of new commodity technology into future CTS, and implications or requirements for future common computing environment (CCE) software development (G1/G2)

~~None~~

FY15

- Participate in CTS-1 ~~TLCC3~~ selection and contract reviews (G1)

FY16

- Accept delivery and install new CTS-1 ~~TLCC3~~ system(s) (G1)

FY17

None

FY18

- Monitor industry hardware and software roadmaps to understand opportunities for cost-effective integration of new commodity technology into future CTS and implications or requirements for future CCE software development (G1, G2)

Advanced Technology Systems (WBS 1.5.4.3)

This level 4 product provides advanced architectures in response to programmatic, computing needs. The scope of this product includes strategic planning, research, development, procurement, testing, integration and deployment, as well as industrial and academic collaborations. Projects and technologies include strategic planning, performance modeling, benchmarking, and procurement and integration coordination. This product also provides market research, and the investigation of advanced architectural concepts and hardware (including node interconnects and machine area networks) via prototype development, deployment, and test bed activities. Also included in this product are cost-effective computers designed to achieve extreme speeds in addressing specific, stockpile-relevant issues through development of enhanced performance codes especially suited to run on the systems.

Sequoia Tri-Lab Advanced Technology Platform (LLNL)

Sequoia is a 20-petaFLOP/s IBM BG/Q advanced technology platform that was sited at LLNL in FY12 with final acceptance planned for early FY13. BG/Q brings many innovations over the previous BG generations, including 16 cores per node, multithreaded cores, a five-dimensional torus interconnect, water cooling, and optical fiber links. The 20-petaFLOP/s system has a staggering 1.6 million processor cores with a total possible 102 million hardware threads all operating simultaneously. This type of parallelism dictates new directions in supercomputing and enters a new regime of the possible physical systems that can be simulated numerically. Codes that are optimized for multi-core and multi-threading will run best on this machine. This platform will be used as a Capability Computing Campaign (CCC) machine for tri-lab stockpile stewardship milestones. Every six months a new CCC process will be run and the next suite of codes will be ushered onto the machine.

Required Capabilities

- R1: Production weapons science calculations at scale (72K nodes and less)
- R2: Production codes as part of CCC process

Gaps

- G1: Effective use of transactional memory
- G2: Effective use of speculative execution
- G3: Efficient use of multi-threaded, multi-core architecture

Five-Year Plan

FY14

- Run two CCC processes
- Continue to investigate optimal performance tuning for specific codes (G1, G2, G3)

FY15

- Run two CCC processes
- Continue to investigate optimal performance tuning for specific codes (G1, G2, G3)

FY16

- Run two CCC processes
- Continue to investigate optimal performance tuning for specific codes (G1, G2, G3)

FY17

- Run two CCC processes
- Continue to investigation optimal performance tuning for specific codes (G1, G2, G3)

FY18

- Retire Sequoia after bringing in 2017 system

Sierra Tri-Lab Advanced Technology System (LLNL)

This is a new project for FY14.

To support the ongoing and coming workload for the stockpile with the necessarily powerful computing systems, the NNSA ASC Program is requesting authorization and funding for a classified UQ- and weapons-science-focused ATS, to be sited and integrated in 2017, that will fill a critical role in support of the Directed Stockpile Work (DSW) mission during the FY18–FY22 period. The Sierra ATS will replace Sequoia and its mission, which will be past its useful lifetime by FY18.

The platform procurement is called "the Sierra ATS procurement," and the implementation project is called the "Sierra ATS integration project." While it will be sited at LLNL, the operation will fall under a proven national user facility paradigm and the system will be available to LANL, LLNL, and SNL.

In keeping with the mission requirement to field an ATS, some portion of the Sierra procurement budget (~10–15%) will be devoted to non-recurring engineering (NRE) work in partnership with the selected vendor.

LLNL is partnering with two Office of Science Labs, Argonne National Laboratory (ANL) and Oak Ridge National Laboratory (ORNL), to acquire three leadership computing systems, one being the Sierra ATS. This collaboration is called CORAL.

Required Capabilities

- R1: Capable of running an uncertainty quantification throughput workload on a suite of integrated weapons performance calculations of relevance and importance to the SSP (primary mission)
- R2: Capable of running a scalable science workload on selected calculations relevant to supporting the predictive capability framework goals (primary mission)
- R3: Capable of running very large benchmark weapon simulations and non-nuclear assessments (secondary mission)
- R4: Accelerate platform innovation by pursuing new technology paths

Gaps

- G1: Effective memory sub-system and sufficient memory capacity
- G2: Low total cost of ownership
- G3: Effective and workable programming model
- G4: Delivery schedule and costs that meets ASC's requirements

Five-Year Plan

FY14

- Release CORAL RFP

- Evaluate CORAL RFP responses and award CORAL non-recurring engineering contract(s) (G1, G2, G3, G4)
- Negotiate and place Sierra build contract
- Complete Sierra CD-1/3a and CD-2
- Provide initial technical coordination and contractual management for CORAL non-recurring engineering and Sierra contracts (G4)

FY15

- Provide technical coordination and contractual management for CORAL NRE and Sierra ATS contracts (G4)
- In collaboration with ANL and ORNL, work with the two NRE contract winners to reduce identified gaps (G1, G2, G3, G4)

FY16

- Provide technical coordination and contractual management for CORAL NRE and the Sierra ATS contracts (G4)
- In collaboration with ANL and ORNL, work with the two NRE contract winners to reduce identified gaps (G1, G2, G3, G4)
- In collaboration with ANL and ORNL, evaluate the NRE results, make the Go/No-Go decision, and negotiate the replacement of target requirements with hard requirements in the CORAL build contracts (G1, G2, G3, G4)
- Complete FY16 facility power modernization

FY17

- Provide technical coordination and contractual management for the Sierra ATS contracts (G4)
- Complete computer room upgrades for siting the Sierra ATS
- Accept delivery of the Sierra ATS

FY18

- Transition the Sierra ATS to the classified network

~~*BlueGene/P and BlueGene/Q Research and Development (LLNL)*~~

~~The BlueGene/P and BlueGene/Q R&D project is a multi-year R&D partnership between NNSA and DOE/SC with IBM on advanced systems. It targets the development and demonstration of hardware and software technologies for 10 petaFLOP/s systems. The 20-petaFLOP/s BlueGene/Q system, Sequoia, was delivered in 2012 to LLNL and the 10-petaFLOP/s BlueGene/Q system, Mira, was delivered to Argonne National Laboratory in 2012. BlueGene/Q system design targets a 20-petaFLOP/s system at the end of the contract.~~

~~This project incorporated requirements from the DOE laboratories, especially Argonne and the NNSA tri-labs, to have input into design choices and system testing for microprocessors, node architectures, and interconnects for BlueGene/Q. The DOE laboratories have provided and continue to provide critical input on software, ensuring appropriate capability and features for the delivered systems.~~

~~***Required Capabilities***~~

~~R1: DOE requirements for advanced systems delivered to IBM for partnership in design choices and systems testing to deliver a successful BlueGene/Q platform~~

~~***Gaps***~~

~~G1: Completion of techniques for optimal use of BG/Q~~

~~***Five-Year Plan***~~

~~FY14~~

- ~~End project; further interactions with IBM on BlueGene/Q technology will be listed under the Sequoia project~~

~~FY15~~

- ~~N/A~~

~~FY16~~

- ~~N/A~~

~~FY17~~

- ~~N/A~~

Hyperion Test Bed (LLNL)

With the extreme demands for capacity computing, the I/O requirements of petascale applications for Sequoia, and the need for improved scientific data management capabilities, it is clearly apparent that emerging breakthrough technologies need to be tested in a large-scale environment such as Hyperion. The Hyperion Test Bed project will work with an expanded set of Hyperion vendor partners in the next phase of the Hyperion project to evaluate innovative node architectures, networks, and alternative storage solutions. Hyperion will continue to be a unique and critical resource for the functionality, performance, stability, and scalability testing of system software.

Required Capabilities

R1: Open, large-scale testing capability for current and future commodity technology systems I/O, computing, and networking technologies

Gaps

G1: Upgraded testing system relevant to future commodity technology systems environment

G2: Evaluate system software for use in future commodity technology systems and high-performance storage

Five-Year Plan

FY14

- Continue to support scalability testing on system software, middleware, storage, and file systems (G2)
- ~~Complete deployment of Hyperion phase 2 server technology refresh, including high performance storage class memory, InfiniBand cluster interconnect, and data intensive capabilities (G1)~~
- Procure and deploy a technology refresh for Hyperion phase 2 servers, including high performance cluster interconnect and potential storage class memory ~~parallel file system storage and storage area network~~ (G1)
- Explore new software models to evaluate the use of high performance storage-class memory and the design impacts of storage-class memory on future system software and hardware architectures (G2)

FY15

- Continue to support scalability testing on system software, middleware, storage, and file systems (G2)
- Complete deployment of new Hyperion phase 2 server technology refresh, including high performance ~~disk and~~ storage class memory, InfiniBand cluster interconnect, and data-intensive capabilities ~~systems into the Hyperion simulation and testing environment~~ (G1)

- Procure ~~and begin deployment of a~~ technology refresh for Hyperion parallel file system storage and storage area network (G1) ~~phase 1 servers, including high performance storage class memory~~
- Explore new software models to evaluate the use of high performance storage class memory and the design impacts of storage class memory on future system software and hardware architectures (G2)

FY16

- Continue to support scalability testing on system software, middleware, storage, and file systems (G2)
- Complete deployment of Hyperion phase 1 server technology refresh, including high performance storage class memory and data intensive capabilities (G1)
- Procure and begin deployment of a technology refresh for the Hyperion phase 2 servers, including high performance storage class memory, cluster interconnect, and data intensive capabilities (G1)
- Explore new software models to evaluate the use of high performance storage class memory and the design impacts of storage class memory on future system software and hardware architectures (G2)

FY17

- Continue to support scalability testing on system software, middleware, storage, and file systems (G2)
- Complete deployment of Hyperion phase 1 ~~2~~ server technology refresh, including high performance storage class memory~~, cluster interconnect,~~ and data intensive capabilities (G1)
- Procure and begin deployment of a technology refresh for Hyperion phase 2 servers, including high performance storage class memory, cluster interconnect, and data intensive capabilities ~~parallel file system storage and storage area network~~ (G1)
- Explore new software models to evaluate the use of high performance storage class memory and the design impacts of storage class memory on future system software and hardware architectures (G2)

FY18

- Continue to support scalability testing on system software, middleware, storage, and file systems (G2)
- Complete deployment of Hyperion phase 2 server technology refresh, including high performance storage class memory, cluster interconnect, and data intensive capabilities (G1)
- Procure technology refresh for Hyperion parallel file system storage and storage area network (G1)

- Explore new software models to evaluate the use of high performance storage class memory and the design impacts of storage class memory on future system software and hardware architectures (G2)

Alliance for Computing at Extreme Scale Trinity Advanced Technology System (SNL, LANL)

The objective of this project is to define requirements and potential system architectures for platforms that meet future ASC programmatic requirements and drivers. The primary activity is to lead the design, acquisition, and plan for deployment of the Trinity Advanced Technology System. The project will coalesce mission requirements, application algorithms, user requirements, and HPC computer industry hardware/software trends into the design process.

Required Capabilities

R1: Architecture, design, and operational support under the ACES MOU

R2: Understanding of tri-lab ASC mission needs and drivers

R3: Understanding of industry roadmaps and the technologies available in the Trinity timeframe

R4: Ability to translate mission needs and vendor capabilities into a platform architecture and associated technical requirements for a successful procurement and deployment process

Gaps

G1: Technologies (hardware and software) may not be available from industry to satisfy mission needs and programmatic usage cases of advanced technology systems

G2: Differences in the platform critical design (CD) processes between NNSA/ASC and Office of Science's NERSC program

Five-Year Plan

FY14

- Evaluate Trinity/NERSC-8 RFP responses and make selection
- Support Trinity acquisition activities
- Complete Trinity CD-2/3b
- ~~Submit Trinity CD2–3 to DOE~~
- Finalize Trinity procurement
- Award Trinity contract
- Commence Trinity development and engineering (D&E) collaborations between selected vendor and the New Mexico Alliance for Computing at Extreme Scale (ACES) ~~R&D collaborations between selected vendor and tri-labs~~
- Update the existing CCC process to accommodate the ATS platforms

FY15

- Complete site preparation for Trinity
- Deliver and install Trinity
- Continue Trinity development and engineering (D&E) ~~R&D~~ collaborations between selected vendor and ACES ~~tri-labs~~

FY16

- Complete Trinity System Integration Readiness Level 2 milestone
- ~~Complete Trinity Production Readiness Level 2 milestone~~
- Enable Trinity operations in support of Capability Computing Campaigns
- Continue Trinity R&D collaborations between selected vendor and tri-labs

FY17

- Run two CCCs
- Complete Trinity Production Readiness Level 2 milestone
- Submit Trinity CD4 to the Department of Energy (DOE) and obtain approval
- Continue Trinity operations in support of CCCs

FY18

- Run two CCCs
- Continue Trinity operations in support of CCCs

Alliance for Computing at Extreme Scale Cielo Capability Computing Platform (SNL, LANL)

The Cielo capability computing platform is a project under the ACES. ACES is a joint collaboration between LANL and SNL defined under an MOU to provide a user facility for capability computing to the NNSA weapons programs in support of stockpile stewardship, to develop requirements and system architecture for ASC capability systems requirements definition, architecture design, procurement, key technology development, systems deployment, operations, and user support.

The architecture and design of Cielo is optimized to provide performance at the full scale of the machine, in support of the NNSA program's most challenging CCCs. This project covers all aspects of the technical, programmatic, and procurement planning for the platform.

Cielo is the primary platform that supports the ASC CCC. In FY10 it replaced the Purple platform at LLNL and is sited at LANL and operated by ACES. Cielo provides 1.37 peak petaFLOP/s with over 140,000 compute cores, and 10 petabytes of storage. Over 6,000 of the cores are dedicated to visualization and data services activities with connections to the SNL and LLNL sites.

Required Capabilities

R1: Architecture, design, and operational support under the ACES MOU

R2: Production, capability-class computing support for the ASC's Capability Computing Campaigns

Gaps

None

Five-Year Plan

FY14

- Complete the Cielo CCC5 and start the CCC6
- Upgrade to Lustre 2.0 on Cielo and attendant platforms
- Continue to run Cielo in production capability mode
- Support production environment of Muzia system as part of the Cielo software quality system (Cielito, Smog, Muzia)
- Provide operations in support of CCCs

FY15

- Provide operations in support of CCCs
- Run two CCCs

FY16

- Provide operations in support of CCCs
- Retire Cielo after successful deployment of Trinity

FY17

None

FY18

None

Next-Generation Computing Technologies (WBS 1.5.4.9)

The Next-Generation Computing Technologies product includes costs for the planning, coordinating, and executing of the next-generation R&D computing technology activities. These activities will prepare the ASC applications and computing environment for the next computing paradigm shift to extreme parallelism, via heterogeneous and/or multi-core nodes.

Next-Generation Computing Enablement (LLNL)

The Next Generation Computing Enablement efforts will help prepare ASC for the post-petascale era, addressing the software environment and platforms. It includes gathering requirements for post-petascale computing and coordinating next generation activities internally and externally. The efforts will enable advanced application work to develop benchmarks for new platforms as well as to adapt codes to the expected new architectures. The software efforts are focused on an architecture that ties together system level software, resource management, development tools, data analysis tools, and programming models, while addressing ASC application requirements. On the hardware side, these efforts include tracking and collaborating on technology innovations. This effort includes interactions with ASCR, vendors, and academia, including planning and technical coordination for vendor contracts. Team members will carry out investigations and co-design activities using test beds and existing technology, making use of proxy applications.

Required Capabilities

R1: Detailed description of post-petascale requirements across software and hardware environment

R2: Understanding challenges the ASC program must overcome to field next-generation systems

R3: Prototypes and simulators to assess the impact on ASC applications

Gaps

G1: Solutions for providing resiliency and power management as scale

G2: Programming models that address the needs of applications that must run at scale on post-petascale platforms

G3: Access to prototypes and simulators

G4: Lack of a unified plan that defines a next-generation software environment at all levels (system, tools, analysis) based on a detailed understanding of requirements

G5: Experiments with proxy applications that inform our understanding of next-generation issues and requirements

Five-Year Plan

FY14

- Develop LLNL plan for software for the Sierra ATS, including programming model, code correctness, power, resilience, and performance tools
- Explore next-generation topics, including characterization of power consumption of key application codes, evaluation of development environment software on advanced architecture test systems, performance characterization, and next-generation resource management

- Participate in planning activities for next-generation computing, including joint meetings with ASCR, meetings and workshops with IC, and interactions with academic collaborators
- Participate in tri-lab effort to produce a next-generation plan that addresses requirements identified through co-design interactions and requirements gathering efforts conducted in the various product areas ~~gathered in FY13 efforts~~ (G4)
- Conduct co-design activities with ASC and ASCR co-design centers and vendors, and research and evaluate next generation technologies (G1, G2)
- Execute LLNL portion of the tri-lab Level 2 milestone to evaluate ~~proxy~~ application~~s~~ performance on advanced architectures ~~Sequoia and test beds~~ (G5)
- ~~Explore post-petascale topics by using test beds and existing petascale technology (G5)~~
- ~~Submit CD2 and CD3 for 2017 ASC Advanced Technology System and provide technical support and coordination for 2017 contract process~~

FY15

- Conduct co-design activities with ASC and ASCR co-design centers and vendors, and research and evaluate next-generation technologies (G1, G2)
- Coordinate and administer PathForward contracts, such as FastForward and DesignForward
- Execute LLNL portion of the tri-lab Level 2 milestone to provide feedback to vendors on key bottlenecks in performance for next-generation technologies (G4)
- Provide technical coordination for 2017 ASC Advanced Technology System contracts
- Identify technology gaps in available systems and procure and deploy prototypes and test beds to allow exploration of how those technologies support ASC applications (G3)

FY16

- Conduct co-design activities with vendors, and research and evaluate next-generation technologies (G1, G2)
- Provide technical coordination for 2017 ASC Advanced Technology System contracts
- Coordinate and administer PathForward contracts, such as FastForward and DesignForward
- Identify technology gaps in available systems and procure and deploy prototypes and test beds to allow exploration of how those technologies support ASC applications (G3)

FY17

- Conduct explorations on 2017 platform (G1, G2, G5)
- Conduct co-design activities with vendors, and research and evaluate next-generation technologies (G1, G2)
- Prepare to site 2017 ASC Advanced Technology System
- Coordinate and administer PathForward contracts, such as FastForward and DesignForward
- Identify technology gaps in available systems and procure and deploy prototypes and test beds to allow exploration of how those technologies support ASC applications (G3)

FY18

- Evaluate next-generation technologies as delivered on the ASC ATS (G4)
- Identify gaps and enhancements needed for technologies in co-design effort with ASC Integrated Codes (G5)
- Coordinate and administer contracts that lay the groundwork for the next generation machines beyond the Sierra ATS
- Identify technology gaps in available systems and procure and deploy prototypes and test beds to allow exploration of how those technologies support ASC applications (G3)

FastForward—Industrial Partnerships for Extreme-Scale Technology Research and Development (LLNL)

The FastForward program is a jointly funded collaboration between DOE Office of Science and NNSA to initiate partnerships with multiple companies to accelerate the R&D of critical technologies needed for extreme scale computing, on the path toward exascale computing. This program is administered by DOE and contracted through Lawrence Livermore National Security, LLC, as part of a seven-lab consortium (Argonne, Lawrence Berkeley, Lawrence Livermore, Los Alamos, Oak Ridge, Pacific Northwest, and Sandia national laboratories).

The first set of five FastForward projects, awarded in summer 2012, included:

- AMD, with focus on processor- and memory-related technologies
- IBM, with focus on memory-related technologies
- Intel, with focus on processor- and memory-related technologies
- Nvidia, with focus on processor and memory-related technologies
- WhamCloud (now owned by Intel), with focus on storage and I/O technologies

Required Capabilities

R1: Commercially available within 5–10 years processor, memory, and storage, and I/O technologies that will maximize energy and concurrency efficiency while increasing the performance, productivity, and reliability of key DOE extreme-scale applications

Gaps

G1: DOE has compelling real-world simulation challenges that will not be met by existing HPC vendor roadmaps in many areas of hardware and software technologies, including processors, memory, and storage, and I/O

Five-Year Plan

FY14

- Provide technical coordination and contractual management for FastForward contracts

FY15

None. Project completed.

Systems Requirements and Planning (LANL)

The Systems Requirements and Planning project covers all aspects of program and procurement planning for current and advanced systems and strategic planning for supporting infrastructure. The major focus is to define requirements and potential system architectures for advanced systems platforms that meet ASC programmatic requirements and drivers. Additionally, this project provides a focus for the various planning efforts. In FY12, this project will focus on the project management of the Cielo system and a FY15 ASC system called Trinity. The focus in this project also includes the execution of DOE Order 413.3. ~~, using the NNSA Office of the Chief Information Officer Project Execution Model for information technology investments.~~

Required Capabilities

R1: Working with vendors and users, identify potential characteristics of platform architectures to meet mission need

Gaps

G1: Define requirements for computing platforms

Five-Year Plan

FY14

- Provide program and project management for computing platforms, including requirements gathering and analysis (G1)
- Participate in site-wide planning for power upgrades for future systems
- Plan infrastructure to support pre-exascale and exascale systems

FY15

- Provide program and project management for computing platforms, including requirements gathering and analysis (G1)
- Begin planning for ATS-3 platform (G1)

FY16

- Provide program and project management for computing platforms, including requirements gathering and analysis (G1)
- Continue planning for ATS-3 platform (G1)

FY17

- Provide program and project management for computing platforms, including requirements gathering and analysis (G1)

FY18

- Provide program and project management for computing platforms, including requirements gathering and analysis (G1)

Next-Generation Computing Technologies (LANL)

~~Co-Design through Mini-Applications (LANL)~~

This project combines the former Co-Design, and Programming Models projects and focuses on the scientific computational environment for the next-generation of computing technologies.

The co-design component of the project leverages other activities at LANL to ~~The Co-Design through Mini-Applications project will~~ build a co-design process through the collaborative creation of patterns, strategies, and abstractions for the implementation and optimization of scientific applications and algorithms on emerging hardware architectures. One aspect of ~~The primary vehicle for~~ this process will be a suite of open source proxy applications, derived from and feeding back into ASC integrated code teams. ASC code teams have informed and continue to ~~will~~ provide the requirements for the study based upon application domains of interest to ASC. Specification documents and reference implementations produced will act as the basis for most of the work in this project, ensuring that it is targeted directly to ASC code needs. ~~The multi-year results of this project will be used as input to new ASC exascale code development efforts.~~

The project will track the development of next-generation hardware architectures and study both computational and data movement patterns represented by the chosen mini-apps developed in conjunction with ASC IC code developers. A major goal of this effort is to inform application developers of methods and best practices that will be necessary for code development on Trinity-like architectures. We will also investigate novel hardware data transformation techniques prototyped using reconfigurable hardware.

Work being done directly with integrated IC teams, in particular the Eulerian Applications Project, will support the evolution of ~~will continue to enhance existing performance and tools, and to evolve~~ current codes towards next-generation architectures by providing computer science expertise on ~~investigating~~ improved mesh data structures, new strategies for AMR, performance improvement, data locality, compressed data structures for materials. In addition, we will explore increased software abstraction, through emerging programming models, and DSLs.

The programming models aspect of the project studies emerging hardware and software trends and their impact on programming abstractions/models, the overall software development tool chain, and run-time systems for scientific-computing environments. LANL's goal is to develop a set of technologies that will assist in the development of the next-generation of application codes as well as extend the lifetime of current codes.

The impact of new architectures on critical IC codes has been categorized at a high level in terms of removing bulk-synchronous communications and increasing levels of concurrency and parallelism. A critical element in meeting these challenges is the adoption of new approaches to programming that reduce the introduction of these characteristics and simplify the programmability of future systems. In FY14, this project will be responsible for a Level 2 milestone that will investigate a potential path forward for reaching these goals by developing and applying these techniques in an implementation of an ASC proxy application and/or a portion of a full IC code. In addition to improving concurrency and reducing synchronization points, LANL will explore approaches that enable interoperability with MPI-based codes to not only minimize the overall impact but also, more importantly, provide a staged migration path for existing codes.

This project also supports a small effort to develop discrete-event simulation techniques to the modeling of advanced file systems such as Panasas and Lustre, and in the future to burst-buffer and other hierarchical and caching storage systems.

Required Capabilities

R1: Provide representative set of proxy applications that explore focused functional capabilities of the ASC code base (for example, unstructured meshes and particle operations)

R2: Provide co-design knowledge base consisting of recommendations and best practices for code development on Trinity and future extreme-scale ~~expected exascale~~ architectures

R3: Working with vendors, explore areas where custom hardware can be applied to applications of interest and develop prototypes for demonstration

R4: Design and demonstration of low-level runtime components, for example, lightweight-thread scheduling and memory affinity on both shared and distributed memory architectures

R5: Language and compiler design and development (including memory and process/thread management), focused on needs of IC and PEM workloads

Gaps

G1: Application design and development practices must support both transition (evolution) and new (revolutionary) code development

G2: Applications will become much more dynamic to support multi-scale, multi-physics, and *in situ* analysis, and will require high-performance data translation and re-formatting as data is exchanged between components of the application

~~G3: Multi-scale and multi-physics codes require data translation and re-formatting as data is exchanged between components of the application~~

G3: Application complexity (for example, dynamic, multi-scale, multi-physics, and resilient) requires system and runtime services to lower development load on application writers

G4: Today's programming models do not fully consider the complexity of memory hierarchies, heterogeneous architectures, and other key requirements for successful application support (for example, resilience and energy)

G5: Today's development tool chain does not fully support emerging complexities of both processor and node-level architectures and the runtime components

~~G5: Heterogeneous hardware is evolving at a rapid rate, and these architectural advances will affect different codes in different ways~~

Five-Year Plan

FY14

- Explore next-generation programming techniques in the context of an ASC integrated code (supports a Level 2 milestone) (G1, G5)
- Characterize existing programming models for application-level resilience and their applicability to ASC science and engineering applications (G4)
- Extend tool-chain support on next-generation architectures for application characterization and analysis (G5)
- Support the development of the programming environment for Trinity (G1)
- Develop proxy applications in support of FastForward efforts and in support of preparing Eulerian Application Project codes for Trinity (G1, G2, G3, G4, G5)
- Implement FPGA-based format translation for next-generation accelerators (G3)
- Extend FileSim capabilities for future file-systems (Lustre or Burst Buffer) (G5)
- Develop proxy applications in support of preparing Lagrangian Application Project codes for Trinity (G1, G2, G3)
- ~~Extend data transformation engine with user programmable functionality via a domain-specific data translation language (G3)~~
- ~~Continue tracking the evolution of next-generation architectures and the associated impact on exemplar applications (G5).~~
- ~~Explore more mini apps and more architectures and extend lessons learned from analysis of mini-apps to realistic applications, as part of tri-lab co-design Level 2 milestone effort (G2, G3, G4)~~

FY15

- Work with ASC developers to fully transition dynamic, multi-physics applications to current heterogeneous hardware (G1, G2, G3, G4)
- Deliver representative set of proxy applications for critical ASC functionality using designed abstractions that compile to multiple hardware architectures (G1, G3)
- Deliver lessons learned and best practices for selected programming models, runtime systems, and architectures for exascale application development (G1, G2, G3)
- Demonstrate custom hardware to dynamically translate multi-scale or multi-physics on-the-fly as it passes between application components (G3)
- Continue to explore and document the advantages and disadvantages of emerging HPC programming models for the needs of the IC and PEM workloads on emerging

architectures; focus on scalability issues, impact of memory hierarchies, and resulting energy implications; support open source-releases (G4)

- Continue to design and implement tool chain capabilities for providing better understanding of code behavior on state-of-the-art architectures, including power usage; support open-source releases (G5)
- Explore and document the advantages and disadvantages of combining node-level and system-wide programming models based on the requirements of the IC and PEM workloads; continue to focus on scalability, energy, and advanced architecture features; support open-source releases (G4)

FY16

- Work with ASC developers to fully transition dynamic, multi-physics applications to current heterogeneous hardware (G1, G2, G3, G4)
- Continue to explore and document the advantages/disadvantages of emerging HPC programming models for IC and PEM workloads on emerging architectures; focus on scalability and resiliency implications and support open-source releases (G4)
- Continue to design and implement tool chain capabilities (G5)
- Work with ASC developers to extend dynamic application to full system, including *in situ* analysis and fault management using services expected to be provided by vendor-supplied system runtime (G1, G2, G3, G4)
- Continue to explore and document the advantages and disadvantages of emerging HPC programming models for the needs of the IC and PEM (G4)

FY17

- Work with ASC developers to extend dynamic application to full system, including *in situ* analysis and fault management using services expected to be provided by vendor-supplied system runtime (G1, G2, G3, G4)
- Continue to explore and document the advantages and disadvantages of emerging HPC programming models for the needs of the IC and PEM (G4)
- Continue to design and implement tool chain capabilities (G5)

FY18

- Work with ASC developers to exploit dynamic applications on extreme-scale computing platforms (G1, G2, G3)
- Continue to explore and document the advantages and disadvantages of emerging HPC programming models for the needs of the IC and PEM (G4)
- Continue to design and implement tool chain capabilities (G5)

~~*Programming Models for the Next-Generation Scientific-Computing Environment (LANL)*~~

~~The goal of this project is to study emerging hardware trends and their impact on programming abstractions/models, the overall software development tool chain, and run-time systems for scientific-computing environments. LANL's goal is to develop a set of technologies that will assist in the development of the next-generation of application codes as well as extend the lifetime of current codes.~~

~~Motivated by the ongoing revolution in computer architectures, this project will experiment with existing techniques and explore new models. It will also support compile- and run-time infrastructures needed to provide the scientific-computing community with the necessary tools to address the challenges of developing software on extreme-scale, highly concurrent systems. Work will be guided by leveraging both existing and evaluation hardware combined with the needs of the IC and PEM workloads, and common computational and communication patterns.~~

~~*Required Capabilities*~~

~~R1: Design of low-level runtime components, for example, lightweight thread scheduling and memory affinity (requires interfacing with operating system activities); this includes both shared and distributed memory perspectives~~

~~R2: Language and compiler design and development (including memory and process/thread management); focused on needs of IC and PEM workloads~~

~~R3: Knowledge of underlying processor/memory architecture capabilities and supporting programming details~~

~~*Gaps*~~

~~G1: Today's programming models do not fully consider the complexity of memory hierarchies, heterogeneous architectures, and other key requirements for successful application support (for example, resilience and energy)~~

~~G2: Today's development tool chain does not fully support emerging complexities of both processor and node-level architectures and the runtime components~~

~~*Five-Year Plan*~~

~~FY14~~

- ~~Continue design and implementation of tool-chain capabilities to better understand code behavior on state-of-the-art architectures; support open source releases (G2)~~
- ~~Explore and document the advantages/disadvantages of node-level programming models combined with system-wide requirements of the IC and PEM workloads; focus on scalability and support open source releases (G1)~~
- ~~Continue to explore and prototype new node-level programming model features and the supporting development tool chain for IC and PEM workloads; focus on code generation/transformations for emerging architectures and support open source~~

~~releases; reduce the complexity of programming while maintaining high performance and maintaining interoperability with existing codes (G1, G2)~~

~~FY15~~

- ~~Continue to explore and document the advantages and disadvantages of emerging HPC programming models for the needs of the IC and PEM workloads on emerging architectures; focus on scalability issues, impact of memory hierarchies, and resulting energy implications; support open source releases (G1)~~
- ~~Continue to design and implement tool chain capabilities for providing better understanding of code behavior on state-of-the-art architectures, including power usage; support open source releases (G2)~~
- ~~Explore and document the advantages and disadvantages of combining node-level and system-wide programming models based on the requirements of the IC and PEM workloads; continue to focus on scalability, energy, and advanced architecture features; support open source releases (G1)~~

~~FY16~~

- ~~Continue to explore and document the advantages/disadvantages of emerging HPC programming models for IC and PEM workloads on emerging architectures; focus on scalability and resiliency implications and support open source releases (G1)~~
- ~~Continue to design and implement tool chain capabilities (G2)~~

~~FY17~~

- ~~Continue to explore and document the advantages and disadvantages of emerging HPC programming models for the needs of the IC and PEM (G1)~~
- ~~Continue to design and implement tool chain capabilities (G2)~~

Co-Design Enablement with Computer Science (SNL)

ASC plans to motivate and influence next-generation, high-performance scientific computing system designs using a co-design methodology. Engineers and scientists from multiple disciplines will work collaboratively to achieve the common goal of useful computers on the anticipated path to exascale computing. On one end of the discipline spectrum are computer hardware engineers (both in industry and at Sandia), and on the other are ASC application developers. Computer scientists play an important role bridging these boundaries of the spectrum by developing capabilities that include: lightweight kernel and runtime system software, programming models, software development environment tools, and architecture-aware algorithms. We will ~~employ computer science information technology and software introspection techniques to assist in the effort, and we will also~~ perform planning and coordination of co-design activities.

Required Capabilities

R1: Critical co-design technical capabilities: HPC architectural simulators, proxy and prototype mini-applications, proxy/proto architectures, advanced architecture test beds and proof-of-concept hardware technology demonstrators

R2: Detailed definition of the post-petascale software environment needs

R3: Detailed definition of the issues that cut across the hardware and software requirements, such as resiliency, power, and changing programming models

Gaps

G1: While SNL has a first-order understanding of the gaps in SNL's path to ~~of~~ exascale (that is, scalability, computational efficiency, energy consumption, programmability, heterogeneity, and resilience), the gaps need to be fleshed out with specifics at all hardware and software levels

G2: SNL needs experiments with proxy and proto mini-applications to gain understanding of our requirements for future platforms

G3: New and existing proxy/proto mini-applications need to be validated against their reference application

G4: Advanced architecture test beds and hardware technology demonstrators will always have limitations; to close these gaps we will need to define and evolve proxy/proto architectures for use in our HPC architectural simulators, driven by our proxy/proto mini-applications

Five-Year Plan

FY14

- Oversee the execution of ~~Execute~~ SNL's portion of the tri-lab Level 2 milestone *Evaluate ~~proxy~~ Application Performance on Advanced Architectures ~~current hardware test beds and/or Sequoia~~* (G2, G4)

- Observe, analyze, and record the future programming environment requirements highlighted by the tri-lab Level 2 milestone *Evaluate Application Performance on Advanced Architectures*

FY15

- Execute SNL's portion of the tri-lab Level 2 milestone *Demonstrate performance and software benefits from proxy apps and architecture explorations of FY14* (G1, G2, G4)
- Summarize and provide lessons learned from prototype mini-Apps to SNL code teams (G3)

FY16

- Provide improved technical requirements for ATS-3 and beyond learned from analysis of proto mini-apps running on prototype architectural simulation models (G1, G2)

  ~~None (expect this project to be replaced with ones that address the identified gaps and requirements)~~

FY17

- Update and summarize lessons learned from prototype mini-apps to SNL code teams (G3)

  ~~None (expect this project to be replaced with ones that address the identified gaps and requirements)~~

FY18

- Provide improved technical requirements for ATS-4 and beyond learned from analysis of proto mini-apps running on prototype architectural simulation models (G1, G2)

Advanced Systems Technology Research and Development (SNL)

~~Initiated in FY12, the Advanced Systems Technology R&D project works to help overcome some of the bottlenecks that limit supercomputing scalability and performance through architecture and software research. In FY13 the focus will continue to be on enabling technologies for co-design.~~

This project will address a critical need for a range of experimental architecture test beds to support path-finding explorations of alternative programming models, architecture-aware algorithms, low-energy runtime and system software, and advanced memory subsystem development. The systems will be used to develop Mantevo proxy applications, enable application performance analysis with Mantevo proxy applications, support the Heterogeneous Computing and Programming Model R&D, the Software and Tools for Scalability and Performance projects, and for SST validation efforts. These test bed systems are made available for "test pilot" users who understand the experimental nature of these test beds. At the present time, it is more important to explore a diverse set of architectural alternatives than to push large scale. Discussions continue with Intel, AMD, IBM, NVIDIA, Micron Technology, and other computer companies regarding ASC interest in obtaining early access to experimental architecture test beds. These partnerships will establish a strong foundation for co-design activities that can influence future hardware designs.

Required Capabilities

R1: Provide a range of advanced system technology experimental architecture platforms for the investigation and advancement of programming model and algorithmic experimentation of ASC applications

Gaps

G1: Understanding scalability characteristics of ASC applications on future architectures

G2: Availability of advanced architecture platforms to validate architectural simulators

G3: Platform availability to investigate advanced power measurement and tuning capabilities

G4: Platform availability to investigate new algorithms for ASC applications

Five-Year Plan

FY14

- Provide platforms to support investigating new programming models, and evaluating compilers and application performance (G1, G2, G3, G4)
- Provide platforms and/or devices for advanced power and energy research and *in situ* application power and energy analysis (G1, G2, G3, G4)

- Provide platform(s) to support Structural Simulation Toolkit (SST) V&V activities (G1, G2, G3, G4)
- Provide platforms for advanced node and platform-level architecture analysis and investigations supporting next-generation platforms (G1, G2, G3, G4)
- ~~Continue to collect validation data for comparison with SST architectural simulations driven by proxy applications (G2)~~
- ~~Characterize ASC application power/energy profiles (G3)~~
- ~~Provide initial performance and application analysis in support of Trinity procurement on candidate advanced architecture platforms (G1)~~
- ~~Provide feedback to industry collaborators on the impact of specific hardware capabilities for future systems (G1, G2, G3, G4)~~

FY15

- Provide platforms to support investigating new programming models, and evaluating compilers and application performance (including existing system technology refresh and network relocation based on program needs) (G1, G2, G3, G4)
- Provide platforms and/or devices for advanced power and energy research and *in situ* application power and energy analysis (G1, G2, G3, G4)
- Provide platforms for advanced node and platform-level architecture analysis and investigations supporting next-generation platforms (G1, G2, G3, G4)
- ~~Provide performance and application analysis in support of Trinity procurement on Trinity specification platform~~
- ~~Provide feedback to industry collaborators on the impact of specific hardware capabilities for future systems (G1, G2, G3, G4)~~

FY16

- Provide platforms to support investigating new programming models, and evaluating compilers and application performance (including existing system technology refresh and network relocation based on program needs) (G1, G2, G3, G4)
- Provide platforms and/or devices for advanced power and energy research and *in situ* application power and energy analysis (G1, G2, G3, G4)
- Provide platforms for advanced node and platform-level architecture analysis and investigations supporting next-generation platforms (G1, G2, G3, G4)
- ~~Provide feedback to industry collaborators on the impact of specific hardware capabilities for future systems (G1, G2, G3, G4)~~

FY17

- Provide platforms to support investigating new programming models, and evaluating compilers and application performance (including existing system technology refresh and network relocation based on program needs) (G1, G2, G3, G4)
- Provide platforms and/or devices for advanced power and energy research and *in situ* application power and energy analysis (G1, G2, G3, G4)
- Provide platforms for advanced node and platform-level architecture analysis and investigations supporting next-generation platforms (G1, G2, G3, G4)
- ~~Provide feedback to industry collaborators on the impact of specific hardware capabilities for future systems (G1, G2, G3, G4)~~

FY18

- Provide platforms to support investigating new programming models, and evaluating compilers and application performance (including existing system technology refresh and network relocation based on program needs) (G1, G2, G3, G4)
- Provide platforms and/or devices for advanced power and energy research and in situ application power and energy analysis (G1, G2, G3, G4)
- Provide platforms for advanced node and platform-level architecture analysis and investigations supporting next-generation platforms (G1, G2, G3, G4)

Application Performance Analysis for Next-Generation Systems (SNL)

The purpose of the Application Performance Analysis project is to develop tools, techniques and methodologies to support the analysis and evaluation of current and next-generation HPC technologies. A primary focus area of the project is to provide leadership for the Mantevo[1] project, and to facilitate the use of Mantevo mini and proxy applications ~~use~~ as a tool. In addition, the project will utilize classic empirical and mathematical performance analysis methods to achieve its goals.

~~A primary activity for FY13 is to perform a study to identify key performance issues of applications executing on emerging technologies.~~ Next-generation computing platforms are expected to present significantly different architectural designs from ~~the~~ previous platforms. ~~several generations~~. In preparation for these changes, the project will explore the potential computing environments from processor ~~core~~, to node, to inter-node. Mini and proxy applications, test beds, simulation capabilities provided by the Structural Simulation Toolkit (SST), abstract machine models, and analytic performance models will be used. The outcome will be a better understanding of the characteristics and capabilities within the context of the computational science and engineering simulations of interest to the ASC program on emerging and future architectures and will inform hardware and software requirements. A primary activity for FY14 is to study and identify key performance issues of applications executing on emerging technologies and, in particular, the Trinity ATS.

Required Capabilities

R1: Expert knowledge of ASC applications, ~~and~~ algorithms, high performance computer architectures, hardware, and software; ability to translate that expert knowledge to definition of mini and proxy applications

R2: ~~Expert knowledge of high performance computer architectures, hardware, and software; ability to translate that knowledge into the design of mini and proxy applications, and~~ Ability to apply mini/proxy applications and classical performance analysis and modeling techniques to evaluate next-generation technologies and platforms

Gaps

G1: Next generation technologies that will have impact on future platforms are not currently available

G2: Current suite of mini and proxy applications are incomplete and do not address all the application capabilities and characteristics

G3: Current simulation and modeling technologies are immature, incomplete, and/or nonexistent

Five-Year Plan

[1] https://software.sandia.gov/mantevo/

FY14

- Assist ~~Work with~~ DOE co-design centers in applying mini-apps and proxy applications ~~and activities that apply to mini and proxy applications~~ (G2)
- Work with the selected Trinity vendor and the tri-labs in the implementation of mini-apps and capability applications for procurement acceptance (G1)
- Develop and deploy (as part of the Mantevo.org suite) an adaptive mesh refinement mini-app (G2, G3)

FY15

- Work with co-design centers and activities that apply to mini and proxy applications (G2)
- Apply mini applications in support of the Trinity procurement (G1)
- Support Level 2 milestone "Using Performance Modeling and Simulation Tools and Techniques to Gauge Key Application Performance Characteristics of the Trinity Platform" (G1)

FY16

- Work with co-design centers and activities that apply to mini and proxy applications (G2)
- Apply mini applications in support of the Trinity procurement (G1)
- Provide support to the ATS-3 design team in defining the role of mini-apps in the procurement evaluation and acceptance phases of the acquisition (G1)

FY17

- Work with co-design centers and activities that apply to mini and proxy applications (G2)
- Provide support to the ATS-3 design team in defining the role of mini-apps in the procurement evaluation and acceptance phases of the acquisition (G1)

FY18

- Work with co-design centers and activities that apply to mini- and proxy-apps (G2)
- Work with the selected ATS-3 vendor and the tri-labs in the implementation of mini-apps and capability applications for procurement acceptance (G1)

Heterogeneous Computing (SNL)

The Heterogeneous Computing project will develop capabilities that facilitate ASC applications' ability to take advantage of heterogeneous architectures with many-core-accelerators, including NVIDIA GPU, Intel MIC, and multicore CPUs. There are clear challenges for coupled, multi-physics-based simulation incorporating unstructured meshes and implicit solution methods. This project will work in tight alignment with co-design efforts and proxy-application development to explore performance on available heterogeneous architecture test beds. Programming models and associated runtime support for portable hybrid parallelism and data locality/placement must be developed to efficiently exploit the diverse set of many-core processors proposed for these architectures. In addition, fine-grained, architecture-aware load balancing of work must be explored and utilized.

These multi-physics coupled applications have a higher level of heterogeneous parallelism that is not currently being exploited. This project will explore SNL's ability to support parallel invocation of coupled applications and efficient management of the shared data. Both physics and engineering applications critical to the nuclear weapons enterprise must be able to take advantage of all levels of parallelism to benefit in the context of exascale computing.

This project will build upon researchers' expertise to improve application performance and portability to the next generation of architectures for scaling applications to a billion-way parallelism. This project has close ties to the heterogeneous architectures test beds in SNL's Advanced System Technology R&D project.

Required Capabilities

R1: A performance-portable programming model that enables a large body of complex, fine grained parallel computational kernels to be efficiently ported to the variety of emerging many-core accelerators, and obtain good performance on these devices with no or negligible code modifications

R2: Proxy-applications that fully exercise this programming model to enable early evaluation of new generations of many-core accelerators

Gaps

G1: Future programming model must integrate fine-grained parallelism and device-polymorphic data locality/placement to obtain performance-portability (the programming model should be implemented via a standard programming language, as opposed to a new or extended programming language)

G2: Lack of programming model/libraries that address memory movement and associated performance impact for complex memory hierarchies with MPI + NUMA + accelerator + accelerator memory regions

G3: Lack of proxy-applications to explore and evaluate advanced many-core, accelerator-based architectures, programming models, hybrid MPI + accelerator parallelism, hybrid heterogeneous MPI + CPU + accelerator parallelism

G4: Lack of a hierarchy of heterogeneous of parallelism starting at the "MPI" level, which is required for effective utilization of exascale parallel resources (for example, a top-level task-based decomposition of multi-physics application where each task is an MPI+accelerator data-parallel computation and accelerator kernels have internal task parallelism)

Five-Year Plan

FY14

- Extend the Kokkos programming model portably to utilize architecture specific capabilities targeting fine-grain parallel performance; for example, Intel Xeon Phi vector instruction units and NVIDIA Kepler shared memory and register shuffle
- Research performance interactions between MPI and Kokkos~~-Array~~ with an emphasis on data movement and overlapping communication with many core computations ~~levels of parallelism, including data movement and effective overlapping of communication and computation~~ (G2)
- Support library and application development teams' prototyping of MPI+Kokkos heterogeneous parallel versions of their codes; incorporate these teams' feedback to improve the programming model
- ~~Document performance guidelines for implementing performance-portable fine-grained parallel algorithms and computational kernels using Kokkos-Array (G1)~~

FY15

- ~~Research MPI + CPU + accelerator heterogeneous parallelism to fully utilize all available compute capabilities (G3)~~
- Collaborate with library and application teams to prototype strategies to migrate their capabilities to heterogeneous computing architectures (G1)
- Research hierarchical, heterogeneous domain decomposition to address NUMA performance concerns and work decomposition in a MPI + CPU + accelerator strategy (G2 ~~G3~~)
- Develop proxy-applications to demonstrate fully utilizing all levels of heterogeneous parallel capabilities within a compute node (G3)

FY16

- Research deeper, heterogeneous parallel work hierarchies for complex multi-physics applications; research programming models to manage the complexity in such hierarchies (G4)

FY17

- Develop proxy-applications with deep heterogeneous parallel work hierarchies (G4)

FY18

- Develop proxy-apps to evaluate programming models for processor-in-memory architectures (G3)

DesignForward—Interconnect Industrial Partnerships for Extreme-Scale Technology Research and Development (LBNL)

This is a new project for FY14.

The DesignForward program is a jointly funded collaboration between DOE Office of Science and NNSA to initiate partnerships with multiple companies to accelerate the R&D of critical technologies needed for extreme scale computing, on the path toward exascale computing. This program is administered by DOE and contracted through Lawrence Berkeley National Lab, as part of a seven-lab consortium (Argonne, Lawrence Berkeley, Lawrence Livermore, Los Alamos, Oak Ridge, Pacific Northwest, and Sandia national laboratories). DesignForward seeks to fund innovative new and/or accelerated R&D of technologies targeted for productization in the 5–10 year timeframe.

The five DesignForward Interconnect projects were awarded in November 2013 to the following companies: AMD, Cray, IBM, Intel, and NVIDIA.

Required Capabilities

R1: Commercially available within 5–10 years the needed interconnect technologies that will maximize energy and concurrency efficiency while increasing the performance, productivity, and reliability of key DOE extreme-scale applications

Gaps

G1: DOE has compelling real-world simulation challenges that will not be met by existing HPC vendor roadmaps in many areas of hardware and software technologies, including interconnect

Five-Year Plan

FY14

- Provide technical coordination and contractual management for DesignForward interconnect contracts

FY15

- Provide technical coordination and contractual management for DesignForward interconnect contracts

FY16

- TBD

FY17

- TBD

FY18

- TBD

System Software and Tools (WBS 1.5.4.4)

This level 4 product provides the system software infrastructure, including the supporting operating system (OS) environments and the integrated tools, to enable the development, optimization, and efficient execution of application codes. The scope of this product includes planning, research, development, integration and initial deployment, continuing product support, and quality and reliability activities, as well as industrial and academic collaborations. Projects and technologies include system-level software addressing optimal delivery of system resources to end-users, such as schedulers, custom device drivers, resource allocation, optimized kernels, system management tools, compilers, debuggers, performance tuning tools, run-time libraries, math libraries, component frameworks, other emerging programming paradigms of importance to scientific code development and application performance analysis.

System Software Environment for Scalable Systems (LLNL)

The System Software Environment for Scalable Systems project provides system software components for all the major platforms at LLNL, research and planning for new systems and future environments, and collaborations with external sources such as the platform partners, especially IBM and Linux vendors. This project covers system software components needed to augment Linux and required proprietary operating systems that function in a manageable, secure, and scalable fashion needed for LLNL ASC platforms.

This project includes work on developing, modifying, and packaging the Tri-Lab Operation Software Stack (TOSS), and developing scalable system management tools to support the OS and interconnect (for example, TOSS and InfiniBand monitoring tools), as well as the resource management environment (Moab and Simple Linux Utility for Resource Management (SLURM)) to queue and schedule code runs across LLNL systems. LLNL uses TOSS on all of its Linux clusters. This project also funds approximately 60 percent of the manpower required to develop, deploy, and maintain TOSS. The funding LLNL receives for its portion of FOUS' TOSS funding accounts for 40 percent of the effort required to develop, deploy, and maintain TOSS. Therefore, TOSS activities and deliverables at LLNL are captured both here and in section 1.5.5.6 of this document.

Required Capabilities

R1: Fully functional cluster software (kernel, Linux distribution, IB stack and related libraries, and resource manager, and cluster-management tools)

R2: Capable of running MPI jobs at scale on Linux capacity clusters

R3: Full lifecycle support, including release management, packaging, QA testing, configuration management, and bug tracking

Gaps

G1: Increasing commodity cluster sizes, increasing core counts, and the ever-present I/O bottleneck will require new system software development activities for efficient operation at ~10,000 node scale

G2: CTS ~~TLCC3~~ hardware support/integration unknown

G3: Integration with Red Hat release schedule and evolutionary changes

G4: Future Linux cluster architectures for large-scale HPC may require support for alternative architectures (for example, Advanced RISC Machine or ARM) ~~accelerators~~

Five-Year Plan

FY14

- Provide ongoing TOSS software development and support (G3)

- Develop/deploy TOSS 2.2 ~~2.X~~ (based on RHEL 6.5 ~~6.X~~) (G3)
- Develop identified system software projects for efficient operation at ~10,000 node scale, including a generic and heterogeneous resource scheduler ~~investigation of new generation of GPU-like processors such as MICs and Fusion~~ (G1, G2, G3)
- ~~Deploy the next generation of the ASC TLCC systems (TLCC3), which includes software integration and testing for the tri-lab environment (G1, G2, G3)~~
- Initiate development of TOSS 3~~.X~~ (based on RHEL 7~~.X~~) (G3)

FY15

- Provide ongoing TOSS software development and support (G3)
- Develop/deploy TOSS 3.X (based on RHEL 7.X) G3)
- Develop/deploy identified system software projects for efficient operation ~10,000 node scale (G1)
- Investigate alternative architectures for commodity Linux clusters (for example, ARM) (G4)

FY16

- Provide ongoing TOSS software development and support (G3)
- Initiate development of TOSS 4 (based on RHEL 8) (G3)
- Develop/deploy identified system software projects for efficient operation ~10,000 node scale (G1, G2, G3)

FY17

- Provide ongoing TOSS software development and support (G3)
- Develop/deploy TOSS 4.X (based on RHEL 8.X) (G3)

FY18

- Provide ongoing TOSS software development and support (G3)

Applications Development Environment and Performance Team (LLNL)

The Applications Development Environment and Performance Team (ADEPT) project provides the code development environment for all major LLNL platforms, supports user productivity, provides research and planning for new tools and future systems, and collaborates with external sources of code development tools. The project works directly with code developers to apply tools to understand and to improve code performance and correctness. The elements of the development environment covered by this project include, but are not limited to, compilers, debuggers, performance and memory tools, interfaces to the parallel environment, and associated run time library work.

Required Capabilities

R1: Robust development environment to support ASC applications running on all major platforms

R2: Tools for researching application code performance and correctness on current and future systems

R3: Compilers and runtimes for all major programming languages and programming models, including C, C++, Fortran, MPI, and OpenMP

R4: Software infrastructure to implement tools and runtimes

R5: Production-level support for checkpoint/restart and parallel I/O

Gaps

G1: We require a diversity of programming models to adequately support future architectures

G2: We lack the resources to field a robust development environment for all possible future capability systems

G3: Promising future technologies may not be mature enough to begin support

G4: New technologies deployed in platforms lead to new user requirements

Five-Year Plan

FY14

- Support and improve the BlueGene/Q development environment for Sequoia/Vulcan ~~Sequoia environment~~
- Continue to support and further enhance the TLCC2 environment
- Assess needs for upcoming CTS-1 environment by investigating new hardware, OS/TOSS, program development, and programming model requirements
- Support tri-lab code teams with performance tuning and debugging support in CCC activities on Sequoia

- Continue the development of new performance analysis, modeling, and code correctness capabilities with a particular focus on scalability
- ~~Prepare for TLCC3 environment and any implied programming model changes (G4)~~

FY15

- Begin scoping of 2017 system requirements for development environment (G3)
- Research and develop resilience strategies for 2015 and future systems
- Provide support for identified emerging programming models (G1, G2)

FY16

- Work with vendor partner identified for 2017 system in co-design for the development environment for the platform (G1, G2)
- Research and develop correctness tools for emerging technologies (G1, G2)
- Support CTS-1 environment

FY17

- After initial system delivery, deploy development environment, test, and validate with user codes
- Continue development of tools for correctness, resilience, performance as gaps are identified (G3)
- Support CTS-1 environment

FY18

- Support user codes on porting, debugging, and performance tuning efforts on the delivered Sierra ATS (G4)
- Support CTS-1 environment
- Support the Sierra ATS environment

High Performance Computing Systems Research (LANL)

HPC systems research is a broad-reaching project focusing on near to long-term research of all the components needed to support a rich environment for very large-scale applications. Systems research bridges the gap between hardware and programming model, and requires tight collaboration in supporting the development of programming models, tools, visualization/analytics, and system software aspect of I/O.

The project is currently focused on resilient system services, soft-error resilience, system support for data-intensive computing, and performance modeling/simulation.

Resilient system services focus on developing a vehicle to investigate resilient, dynamic, distributed, scalable services for large-scale systems and providing an interface to programming models so that ASC applications can access these features on current and future hardware. Current activities include investigation of distributed systems software for job launch and monitoring. An imminent challenge for extreme-scale HPC systems is the issue of power limits and rapidly varying demands on the grid. Techniques for power-capping HPC systems will be investigated.

Investigations of soft-error resilience will continue work on ~~focus on creating~~ an accelerated testing environment for soft error profiling using a virtual machine (VM) approach to inject faults while actively running real ASC codes. Hardware reliability in HPC systems remains a challenge to characterize. Statistical studies of reliability data from a variety of production systems will be extended, and models of DRAM reliability will be developed. Reliability of non-volatile storage will be studied in support of future hierarchical storage systems (for example, burst buffer architectures).

The performance simulation investigation is currently focused on application I/O patterns and transparent checkpointing, and on lock performance in hybrid MPI/thread programs from an OS perspective. ~~used on parallel file systems under realistic application loads.~~

Required Capabilities

R1: Distributed systems software expertise

R2: OS support for future HPC applications and runtimes

R3: Resilience of software and systems

R4: OS support for future HPC applications and runtimes

Gaps

G1: Large-scale systems are exceeding the limits that centralized system software and services can support; we look to distributed systems research to enable scaling of system software and services in a resilient way to future ASC systems

G2: Resilience of applications and systems is a cross-cutting issue; we investigate profiling of applications for susceptibility to soft-errors in systems and applications

G3: Lack of adequate characterization and models of hardware reliability in HPC systems

~~G3: Development of a simulator for parallel file systems will provide predictive capability for design of file systems for future large-scale systems~~

G4: Inadequate power-management techniques for next-generation computing platforms at extreme-scale

Five-Year Plan

FY14

- Develop distributed software techniques for HPC job launch and monitoring (G1, G2)
- Model power-capping for TLCC-class systems with priority queues (G4)
- Analyze production system statistics related to reliability (G3)
- Extend Chipkill DRAM reliability models (G3)
- Develop a model for predicting DRAM reliability on current and next-generation machines (G3)
- Explore sub-domain indexing in data-intensive system software for N to M restarts (G1, G2)
- Characterize performance of advanced HPC system software, including transparent checkpointing, thread-MPI synchronization, and other programmatically driven needs (G1)
- Investigate performance and scalability of next-generation interconnect topologies
- Conduct a detailed statistical analysis of reliability data from HPC systems at various HPC facilities (G3)
- ~~Extend resilient system services to burst-buffer management (G1)~~
- ~~Investigate optimization of open-flow for HPC protocols and scalable system service protocols (G1)~~
- ~~Extend the toolset for soft error susceptibility (G2)~~
- ~~Develop solutions for Trinity scalability issues as Trinity test beds are available (G1)~~
- ~~OS modifications for *in-situ* processing (G1)~~

FY15

- Investigate adaptive runtimes, supported via OS for interconnected many-core systems (G1, G2)
- Explore power-capping techniques for HPC at scale (G4)
- Refine models of memory system reliability for next-generation HPC systems (G3)

FY16

- For self-healing OS, integrate system services to respond to hardware and software system errors and performance issues (G1, G2, G3)

FY17

- For global OS, integrate node OS and system services into cohesive distributed system (G1, G2, G3, G4)
- Continue to characterize and model reliability of emerging technologies for HPC (G3)
- Provide OS and runtime support for emerging programming models for next-generation systems (G1, G2, G3)

FY18

- Improve performance and scalability of global OS and system services (G1, G2, G3, G4)
- Provide OS and runtime support for emerging programming models for next-generation systems (G1, G2, G3)
- Continue to characterize and model reliability of emerging technologies for HPC (G3)
- Provide OS and runtime support for emerging programming models for next-generation systems (G1, G2, G3)

Test Beds (LANL)

TBD

Required Capabilities

R1: TBD

Gaps

G1: TBD

Five-Year Plan

FY14

- Manage computer test beds for CSSE and IC project use
- Upgrade laboratory test bed air-conditioning units
- Refresh technology in Darwin test bed system

FY15

- TBD

FY16

- TBD

FY17

- TBD

FY18

- TBD

~~*Application Readiness (LANL)*~~

~~The Application Readiness project addresses issues with an application's production-run readiness on current and incoming computing systems at LANL. Working with subsystem teams such as systems management, file systems and I/O, archive, and tools, the Application Readiness team identifies causes of unexpected behavior and deploys fixes in production. The project goal is that system users are able to make productive use of the systems with their applications to solve their problems.~~

~~The project provides production problem solving (create small problem reproducers, identify cause, consult with the relevant technical experts to find a solution, and verify the deployed solution), periodic stress testing/regression of production machines, new software version regression testing, system configuration verification and software stack deployment with real user applications and metrics, and analysis/profiling.~~

~~***Required Capabilities***~~

~~R1: All LANL and ACES workload performing at high efficiency and reliability~~

~~R2: Difficult application problems rapidly diagnosed and resolved~~

~~R3: System problems that impact applications rapidly identified with full supporting data~~

~~R4: New and upgraded systems need to have their application functional; performance and reliability capabilities rapidly assessed~~

~~***Gaps***~~

~~G1: HPC systems are becoming larger, which stresses the scalability of applications~~

~~G2: HPC systems are becoming more complex with a diversity of technologies and programming models, which stresses the capabilities of diagnostic tools and techniques~~

~~G3: Advancing size and complexity of HPC systems makes rapid characterization of application performance difficult~~

~~***Five-Year Plan***~~

~~FY14~~

- ~~Continue to support users of Luna, Moonlight, Cielo Lustre FS, and other recently integrated (sub) systems by tackling hard-to-diagnose problems, typically involving the interaction of applications with multiple aspects of the computational environment~~
- ~~Continue to work with other projects, including Co-Design through Mini-Apps, Programming Models, and Code Strategies for Emerging Platforms to assist code development teams with migration towards the use of hybrid programming to exploit the GPU-based elements of a new TLCC2~~
- ~~Deploy Application Observation tools with key applications~~

- ~~Integrate Application Observation tools with new system assessment application suite~~

~~FY15~~

- ~~Continue to support applications on existing and new systems~~
- ~~Support applications being ported to new architectures on Trinity prototype systems~~
- ~~Migrate Application Observation tools to support new architectures and programming models~~

~~FY16~~

- ~~Support users of Trinity (and other LANL and ACES systems) by tackling hard-to-diagnose problems, typically involving the interaction of applications with multiple aspects of the computational environment~~
- ~~Deploy Application Observation tools to Trinity~~

~~FY17~~

- ~~Continue support for users of Trinity (and other LANL and ACES systems) by tackling hard-to-diagnose problems, typically involving the interaction of applications with multiple aspects of the computational environment~~

~~*Software Support (LANL)*~~

~~The Software Support project works to establish a strong development and analysis tool capability for current and next-generation HPC platforms, including parallel capabilities. It is focused on working with the HPC tool community and vendors to identify, plan, and integrate tools into production environments and establish a solid support structure. The project supports the incremental improvement of tools driven by ASC strategic needs, including cross-laboratory partnerships and external collaborations that focus on performance tools required for programming model support.~~

~~Capabilities include development of a strategic plan for tools based on current and next generation platform planning, integration with HPC community tool development efforts, and tool development and production integration capability.~~

~~***Required Capabilities***~~

~~R1: Strategic/support plans and collaborations to continue to increase scalability and support of Programming Environment and Analysis Tools~~

~~R2: Strong MPI development and support~~

~~R3: Software support and development capability strategy plan~~

~~***Gaps***~~

~~G1: Programming models and HPC architectures are changing rapidly with limited tool support~~

~~G2: Increased scale of HPC systems pushing the limits of what software tools can provide~~

~~***Five-Year Plan***~~

~~FY14~~

- ~~Provide MPI development and user support for optimization (Open MPI) (G1, G2)~~
- ~~Continue interaction with community and vendors to increase MPI and tool scale capability (G1, G2)~~
- ~~Integrate monitoring environment to initial performance analysis tool through CBTF environment (G1, G2)~~
- ~~Provide support for MPI+X strategy on Trinity and successor platforms (G1)~~

~~FY15~~

- ~~Provide MPI development and user support for optimization (Open MPI) (G1, G2)~~
- ~~Continue interaction with community and vendors to increase MPI and tool scale capability (G1, G2)~~
- ~~Support infrastructure for tool deployment and support for platform architecture and programming models being utilized (G1, G2)~~

~~FY16~~

- ~~Provide MPI development and user support for optimization (Open MPI) (G1, G2)~~
- ~~Continue interaction with community and vendors to increase MPI and tool scale capability (G1, G2)~~
- ~~Support infrastructure for tool deployment and support for platform architecture and programming models being utilized (G1, G2)~~

~~FY17~~

- ~~Provide MPI development and user support for optimization (Open MPI) (G1, G2)~~
- ~~Continue interaction with community and vendors to increase MPI and tool scale capability (G1, G2)~~
- ~~Support infrastructure for tool deployment and support for platform architecture and programming models being utilized (G1, G2)~~

Software and Tools for Scalability and Performance (SNL)

The Software and Tools for Scalability and Performance project supports system software R&D to address scalability and efficiency of future computational systems in multiple dimensions. An important aspect is providing lightweight services and functionality that do not compromise scalability and therefore performance. The focus starting in FY13 will be on three emerging critical areas for HPC systems, which will enhance efficiency, performance and scalability of applications on future HPC systems:

- Power has become a first-order design constraint for future supercomputers. SNL will expand upon work in data collection and tuning techniques that provided new insight into understanding power requirements and affecting power use of ASC applications.
- Dynamic Shared Library (DSL) support on HPC systems has also become high profile in recent years. Supporting DSLs in a scalable manner on future HPC systems is critical for many ASC applications.
- Previous work with virtualization has shown promise in the area of HPC. Virtualization will be leveraged to provide insights into application runtime characteristics and where optimization efforts would be best targeted.

As a long-term goal, SNL plans to integrate these targeted efforts with previous successes in lightweight kernel (Kitten) and high performance network stack (Portals) development with a production HPC computing stack. While this is a significant development effort, the long-term benefits are many:

- Risk mitigation against vendor provided software failure
- Proof-of-concept demonstration of scalable or efficiency enhancements at scale
- Setting the standard for vendor-delivered scalable software stacks
- Enabling scalable research in both systems software and other areas dependent on scalable software features
- DSL large scale testing and optimization

This effort will necessarily be accomplished in conjunction with the acquisition of the next ACES Advanced Technology System (Trinity).

Required Capabilities

R1: New software features and capabilities are required to maintain and enhance scalability, productivity, and efficiency of applications on advanced architecture platforms

Gaps

G1: Understanding power efficiency on large-scale, advanced architecture HPC systems

G2: Dynamic Shared Library support at scale on HPC systems

G3: Virtualization capabilities on production HPC hardware

G4: Lightweight kernel development for HPC

Five-Year Plan

FY14

- Define power/energy API at all ~~necessary~~ levels identified in scope of *Power API Use Case* document ~~: Reliability Availability and Serviceability (RAS), operating system, runtime, and application level~~ (G1)
- ~~Perform benchmark analysis of scalable virtual machine environment running on a production ASC platform (G3)~~
- Enable two-way communication between LWK and dynamic adaptive runtime system (G4)
- Prototype operating system/hardware interface portion of Power API (G1)
- Prototype hardware measurement portion of Power API using PowerInsight (G1)

FY15

- Perform benchmark analysis of node virtualization layer running on a production ASC platform (G3)
- Integrate power/energy measurement and API on Trinity early spec platform (G1)
- ~~Develop system-directed virtual machine migration based on dynamic runtime behavior and resource availability (G3)~~
- Evaluate effectiveness of LWK and dynamic adaptive runtime system combination for ASC workloads (G4)

FY16

- Demonstrate application portions of power/energy measurement and API on ACES Trinity platform (G1)
- Characterize and demonstrate increased application energy efficiency on Trinity (G1)
- Port scalable virtual machine environment to Trinity (G3)

FY17

- Perform benchmark analysis of node virtualization layer ~~scalable virtual machine environment~~ on Trinity (G3)
- Deploy production capability of power/energy measurement and API on ACES Trinity (G1)

FY18

- Evaluate LWK with production workloads at large scale on Trinity platform (G3)

Resilience (SNL)

Resilience is a critical issue for future extreme-scale computing systems. Resilience in this context refers to the ability of applications to run correctly to completion even in the face of failures in hardware and system software. The Resilience project's goal is to develop scalable techniques, mechanisms, and enhancements to application runtime and system software to reduce the impact of failures to enable successful application execution at extreme scales. The work comprises two components: 1) improving our understanding of the fault phenomena (fault modeling and root cause analysis), and 2) developing strategies and tools to respond to faults/degradation to permit application execution amid faults (transient, permanent, and undetected). This work also involves addressing the integrated resilience problem "framing" to develop a full-system software architectural paradigm to manage faults on extreme-scale systems.

Required Capabilities

R1: The ability to rapidly assess root-cause of system faults and inform the OS, runtime system, and applications so they can contain and manage the faults

R2: A comprehensive fault-management framework with standardized APIs at the runtime system and application levels

R3: The ability to define persistent data object stores that will support scalable fault management at the application level

R4: The ability for application programmers to declare specific segments of code to execute reliably with a designated level of assurance

R5: Inherently robust problem formulations (for example, PDE stencil), such that the applications can proceed correctly in the event of silent (that is, undetected) errors

R6: Runtime support for task-based programming models and the ability to manage fail-overs, task rescheduling, and dynamic load balancing amid system faults and degradation

Gaps

G1: Lack of comprehensive model of system faults necessary to understand the root causes of the failures and required for us to develop system-wide fault management approaches and systems

G2: Lack of a comprehensive fault management framework and the associated set system abstractions defining computing models ~~a resilience architecture~~

G3: Need for scalable alternatives to global checkpoint restart that can be exercised at the application and/or runtime level

G4: Ability to specify and execute accordingly differing degrees of reliability for particular code segments within an application

G5: Mechanisms for both improved capability to detect so-called silent errors and to reframe applications and algorithms to be robust in the presence of the most common silent errors (for example, undetected bit flips)

G6: Mechanisms to support highly dynamic application task scheduling, remapping, migration, and process/data cloning (redundancy) execution models

Five-Year Plan

FY14

- Model the relationships between error messages, failed jobs, and likely root causes based on data from production systems but relevant to advanced architectures (G1)
- ~~Generalize fault model to other architectures~~
- ~~Quantify anomaly detection algorithms (for example, the effectiveness of "nodeinfo") effectiveness for discovering new fault types on production systems (G1)~~
- Explore parallel scalability of robust stencil approach for silent-error tolerance (G5)
- Investigate optimized algorithmic approaches for physics simulations that damp bit-flips more efficiently by operating in concert with specific physical dynamics ~~that reduce the overhead of bit-flip damping~~ (G5)
- Demonstrate LFLR resilient computing model for PDE mini-app in simulated scalable environment with MPI process loss (G2, G3)
- Define prototype LFLR persistent storage API and semantics (G2, G3)
- ~~Demonstrate a dynamic scheduling and remapping scheme for task-based application, including analysis of system performance on the SST simulator (G6)~~

FY15

- Assess the effect of HPC architecture on how silent errors manifest in applications, to develop adaptive HPC design strategies to support resilience (G4, G5)
- Demonstrate vertically integrated application resilience, including prototype reliable code segment specification and integrated fault models and detectors with system response mechanisms (G1, G2, G3, G4, G5, G6)
- Implement LFLR persistent storage API on leadership-class system (G2, G3)
- Define prototype selective unreliability API and semantics (G2, G4, G5)

FY16

- Demonstrate a representative robust simulation application at scale on a production system and evaluate its performance and silent-error tolerance (G5)
- Demonstrate vertically integrated application resilience on a production system of an ASC IC application or library component (G2, G3, G4, G5, G6)

FY17

- Impact co-design of next-generation HPC architectures to optimize support for efficient resilience mechanisms (G2, G3, G4, G5, G6)
- Specify APIs, requirements and execution model for integrated resilience capabilities, ~~deploy~~ initial production versions of core capabilities (G2, G3, G4, G5, G6)

FY18

- Deploy integrated resilience capabilities in production for ASC codes on extreme-scale systems (G1, G2, G3, G4, G5, G6)

System Simulation and Computer Science (SNL)

The Structural Simulation Toolkit (SST) is a suite of tools enabling multi-scale computer architecture simulation to meet the needs of HPC software/hardware co-design. The SST consists of a core set of components that enable parallel discrete-event simulation; high-fidelity networking, memory, and processor components; and coarse-grained simulation components that capture essential elements of machine performance with low computational cost. Future HPC systems and the applications designed to utilize them are impacted by a variety of considerations, including scalability of applications, ease-of-programming, memory and network latencies becoming more imbalanced relative to computation rates, data corruption and its propagation, frequency of interrupts, power consumption, and overall machine cost. SST is designed to allow each of these parameters to be explored, permitting the consideration of a broad space of potential architectural and application/algorithmic designs. The goal is for the SST components to be extended and enhanced by a community of simulator developers, including academic, industrial, and government partners. An even larger community is expected to be the users of SST, including algorithm developers, architecture designers, and procurement team members.

Required Capabilities

R1: Ability to understand potential performance of large-scale HPC systems before they are built

Gaps

G1: Need scalable and fast hardware simulation to enable prediction of application performance on new hardware

G2: Current model validation efforts of hardware simulators are inadequate

G3: Need a complete suite of integrated hardware models

Five-Year Plan

FY14

- Perform large-scale macro simulations for design-space exploration of next-generation ~~exascale~~ machines and applications (G1)
- ~~Explore emerging programming and execution models at macro scale (G1)~~
- Validate SST/macro against more architectures/machines to demonstrate the breadth of capability and improve the quality of existing models (G2)
- Work towards automating workflows around SST/macro, including creating skeletons of existing applications and uncertainty quantification (G1)
- Publish to community a library of DUMPI traces taken on different machines for different proxy/mini-apps (G1)

- Report results of validating the predictive capabilities of two micro-scale research questions ~~hardware modules~~ (G2)
- Verify newly introduced SST/Micro threading code via experimentation with Mantevo applications using OpenMP (G3)
- Use scheduler and power models to assess the performance impact on energy of proposed cluster-level task allocation strategies integrated with a power estimation methodology (G3)
- Continue ongoing usability and integration efforts (G1)

FY15

- Report modeling done on Trinity components (G1)
- Provide a pre-configured, validated, "plausible machine" with which application developers can run their code (G1, G2, G3)
- Provide a validated power and energy consumption modeling mechanism (G2)

FY16

- Demonstrate a use case for mixed model (macro/micro) simulation (G1)
- Create an SST/Micro model that can aid in the analysis of reliability of for example, processor components and memory buses (G1)

FY17

- Down select hetero-core models to the one or two that have gained traction in the recent years (G2)
- Demonstrate a macro simulation of a proxy application that was migrated from an existing form (MPI-everywhere) to an exascale-compatible programming/execution model (G1)

FY18

- Using SST/Micro, analyze and report on the tradeoffs of heterogeneous big-core and little-core processors (G1)

Scalable, Fault-Resilient Programming Models (SNL)

ASC has identified several key software development challenges on the road to exascale. While the several-orders-of-magnitude increase in parallelism is the most commonly cited of those, hurdles include drastically shortened mean times to interrupt, increased imbalance between computational capacity and I/O capabilities, silent errors, and complex hardware architectures. Various programming and execution models have been proposed to address these issues, but before ASC applications can use these, detailed insight will be required into the level of porting effort, scalability, and fault tolerance characteristics of these approaches.

Current efforts on architecture simulation, performance prediction and analysis, and mini-application analogues of full-scale ASC codes will be used to identify required algorithmic changes and to quantify the performance and fault resilience characteristics of proposed exascale programming models. Efforts will focus on both on-node execution models required to support upcoming heterogeneous architectures in an "MPI+X" parallel environment, and on alternate software development approaches with potential to support greater levels of parallelism and fault resilience than the cooperative message passing SPMD approach commonly associated with MPI.

Required Capabilities

R1: ~~A~~ We need a performant, portable, scalable, fault-tolerant programming model that decreases the application programming development workload while increasing the utilization of the application code at extreme scale

R2: ~~Ways~~ We need ways to compose applications so that any new programming models are minimally disruptive and generally decrease ~~increase~~, not increase ~~decrease~~, the code rewrite load on the application developers

R3: We need the runtime and system level components required to enable fault-tolerant programming models and in particular, task-based, data-flow execution models

Gaps

G1: Current bulk-synchronous (SPMD) programming models are neither inherently fault tolerant nor do they work effectively at the levels of concurrency required on future extreme-scale architectures. We need alternative programming models that are asynchronous, scalable, and inherently support fault tolerance and dynamic load balancing.

G2: We need ways to lower the barriers to using these new programming models, including supporting programming and debugging tools, and guidelines and methodologies on how to compose applications comprised of part legacy programming models along with new, different programming models (for example, bulk-synchronous SPMD legacy code using components/libraries developed with asynchronous, task-DAG data-driven programming models).

G3: We need the runtime and OS components that support task-DAG, data-driven asynchronous programming models.

Five-Year Plan

FY14

- Explore design tradeoffs for scheduling within a resilient, asynchronous, distributed-memory task-DAG runtime using SST/macro simulations (G1, G3)
- Develop an SST/macro implementation of the distributed task-DAG runtime and use it to study the performance and fault-resilience limits of the task-DAG approach for a port of mini-FE on an advanced technology architecture (G1, G3)
- Based on the results of SST/macro experiments, begin implementing components of a resilient, asynchronous, distributed-memory task-DAG runtime to be used for SNL's FY15 Level 2 milestone (G1, G3)
- ~~Demonstrate a vertically integrated, asynchronous, task-DAG implementation for parallel mechanical contact application on a realistic problem at scale (G1, G2, G3)~~
- ~~Study performance and limits of the task-DAG approach to broach class of fault on various capability-class system architectures, and develop APIs to support future development of asynchronous, task-DAG approaches (G1, G2, G3)~~
- ~~Develop optimized dynamic scheduling and replanning methods for asynchronous, task-DAG programming models in support of the contact application demonstration on SST and a capability class system (G1, G3)~~

FY15

- ~~Explore how to phase task-DAG support into legacy codes without requiring a total rewrite; examine issues of composition of applications/libraries where both MPI+X and task-DAG programming models are present (G1, G2)~~
- Demonstrate vertically integrated asynchronous task-DAG programming model on a realistic IC code or library component on both SST and a capability class machine~~;~~ ~~address composition of the task-DAG approach in conjunction with a legacy (MPI+X) code~~ (G1, ~~G2,~~ G3)
- Develop optimized dynamic scheduling and replanning methods for asynchronous, task-DAG programming models in support of the application demonstration on SST and a capability class system (G1, G3)

FY16

- Explore how to phase task-DAG support into legacy codes without requiring a total rewrite; examine issues of composition of applications/libraries where both MPI+X and task-DAG programming models are present (G1, G2)
- Demonstrate vertically integrated asynchronous task-DAG programming model on parallel mechanical contact application on both SST and a capability class machine;

this demonstration should address composition of the task-DAG approach in conjunction with a legacy (MPI+X) code (G1, G2, G3)

- ~~Develop general guidelines, examples, runtime elements, tools, and documentation for widespread use of asynchronous task-DAG programming models on capability class systems (G1, G2, G3)~~

FY17

- Develop general guidelines, examples, runtime elements, tools, and documentation for widespread use of asynchronous task-DAG programming models on capability class systems (G1, G2, G3)
- ~~Develop production ready implementations, examples, runtime elements, tools, and documentation for widespread use of asynchronous task-DAG programming models on capability class systems (G1, G2, G3)~~

FY18

- Develop general guidelines, examples, runtime elements, tools, and documentation for widespread use of asynchronous task-DAG programming models on advanced technology class systems (G1, G2, G3)

Input/Output, Storage Systems, and Networking (WBS 1.5.4.5)

This level 4 product provides I/O (data transfer) storage infrastructure in balance with all platforms and consistent with integrated system architecture plans. The procurement of all supporting subsystems, data transfer, storage systems, and infrastructures occurs through this product. The scope of this product includes planning, research, development, procurement, hardware maintenance, integration and deployment, continuing product support, quality and reliability activities, as well as industrial and academic collaborations. Projects and technologies include high-performance parallel file systems, hierarchical storage management systems, storage-area-networks, network-attached storage (NAS), and high-performance storage system (HPSS) or future hierarchical storage management system disks, tape, robotics, servers, and media. This product also includes relevant prototype deployment and test bed activities. Projects and technologies in the advanced networking and interconnect areas include networking and interconnect architectures, emerging networking hardware technologies and communication protocols, network performance/security monitoring/analysis tools, and high performance encryption and security technologies.

Archive Storage (LLNL)

The Archival Storage project provides long-term, high-performance, archival storage services to ASC customers. This includes a collaborative software development effort (currently, HPSS) between the tri-labs, Oak Ridge National Laboratory, Lawrence Berkeley National Laboratory, and IBM, as well as deployment and support of archival storage software and interfaces for tri-lab ASC customers on unclassified and classified networks. It includes the selection, procurement, deployment, support, and maintenance of archival storage hardware and media, ongoing technology refresh, and data stewardship. HPSS provides scalable, parallel, archival storage interfaces and services to the tri-labs.[2]

A world-class array of hardware is integrated beneath HPSS, supplying the performance necessary to offload ASC platforms, thereby increasing computation. This includes disk arrays, robotic tape subsystems, servers, SANs, networks, and petabytes of tape media, enabling high-speed parallel transfers into a virtually unlimited data store.

Required Capabilities

R1: Cost-effective, long-term, high-performance archival storage services for ASC customers

Gaps

G1: Tape drive vendor roadmaps currently provide adequate bandwidth increases required for cost-effective archives in pre-exascale HPC environments. To ensure long-term viability and cost-effectiveness of archives at exascale, tape drive bandwidths need to radically increase, operational approaches need to be adjusted (for example, RAIT), or a disruptive archive technology needs to be identified to replace tape.

G2: Scalability of archive metadata performance and capacity is limited by the current single-instance database design of HPSS. Future versions must focus on providing partitioned metadata initially, and potentially multiple metadata servers to increase the total number of metadata transactions.

G3: As HPC environments grow more complex, any number of data path components can silently introduce data corruption. Data checksum algorithms can verify data integrity, but this feature is currently missing from HPSS. To ensure long-term data correctness, end-to-end data checksum features should be developed in HPSS and deployed in archival storage environments.

Five-Year Plan

FY14

- Continue ongoing HPSS software development and support, with focus on development and testing of ~~Test and release~~ HPSS 7.P ~~(Panda)~~, ~~featuring~~ which

[2] See http://www.hpss-collaboration.org/index.shtml.

features partitioned metadata and is currently targeted for General Availability in CY2015 as HPSS 7.5 (G2)

- ~~Evaluate, select, procure, and deploy upgrades to enterprise tape drive environment for increased archive capacity and bandwidth (G1)~~
- Deploy HPSS 7.4.x with conversion to native 64-bit architecture, support for IPv6, dynamic update of devices/drives, RAIT, and enhanced repack of legacy small files into aggregates ~~support for RAIT~~ (G1)
- Evaluate and potentially procure and deploy upgrades to enterprise tape drive environment for increased archive capacity and bandwidth, and begin repack of 8-year old T10K Gen1 media (1-TB native) to T10K Gen2 media (5-TB native) to minimize data loss due to aging media and to reclaim slot capacity in libraries
- Develop and deploy system-based software to automatically failover dual-homed HPSS disk cache devices to alternate disk movers to increase archive availability
- Deploy X86-based platforms and Oracle's Automated Cartridge System Library Software (ACSLS) 8.x for Oracle-based tape libraries
- Provide ongoing support of currently deployed archival storage systems, including selection, deployment, support, and maintenance of all archival storage hardware and media, customer and interface support, ongoing tech refresh, and data stewardship

FY15

- Develop future version of HPSS featuring checksums and distributed servers and metadata for increased scalability of metadata performance and capacity (G2, G3)
- Evaluate, select, procure, and deploy hardware upgrades to mid-range tape drive environment for increased archive capacity and bandwidth (G1)
- Deploy network upgrades from 10Gbe to 40Gbe for increased archive performance

FY16

- Finalize, test, and release future version of HPSS featuring checksums and distributed servers and metadata for increased scalability of archive metadata performance (G2, G3)
- Evaluate, select, procure, and deploy hardware upgrades to HPSS core server, metadata, and ~~aging~~ disk subsystems for increased archive metadata performance, capacity, and bandwidth (G1)

FY17

- Develop future version of HPSS featuring additional metadata scalability features in preparation for exascale computing needs (G2)
- Deploy future version of HPSS featuring checksums and distributed servers and metadata for increased scalability of archive metadata performance (G2, G3)

- Evaluate, select, procure, and deploy upgrades to enterprise tape drive environment for increased archive capacity and bandwidth (G1)

FY18

- Finalize, test, and release future version of HPSS featuring additional metadata scalability features in preparation for exascale computing needs (G2)
- Deploy network upgrades from 40Gbe to 100Gbe for increased archive performance

Parallel and Network File Systems (LLNL)

The Parallel and Network File Systems (NFS) project provides for the development, testing (feature, capability, performance, and acceptance), procurement, integration, and ongoing support of various file system technologies and interfaces necessary for the efficient and effective use of ASC high-performance platforms. Included are the continuing development and support of Lustre as a fully featured file system for the range of ASC platforms, the deployment and support of ubiquitous NAS services for home, project, and scratch space, and the I/O support of various programming interfaces for parallel I/O.

This project deploys and supports Lustre file systems for ASC platforms as well as high-availability NAS file systems for home and project space, and scratch space for serial capacity clusters. It actively works with the OpenSFS Lustre development community to add Lustre file system scalability and reliability enhancements required by ASC platforms. The file system up through the programming interfaces are supported to help developers of applications use parallel I/O effectively.

Required Capabilities

R1: Provide reliable and available high-speed parallel file system access in support of ASC platforms

R2: Maintain optimal Lustre and NFS functionality across Center machines for local and remote users

R3: Develop and deploy Lustre HPC-focused software releases in concert with broader Lustre community

R4: Deploy new file systems consistent with platform requirements

R5: Work with users and application developers on developing and improving I/O access mechanisms

Gaps

G1: Lustre file system performance limitations ~~underpinnings limit file system performance and scalability~~

G2: Lustre file system availability and reliability ~~suffer at scale in the areas of file system recovery, data integrity, and integrity checking and negatively impact users following off-normal events~~

G3: Parallel file system TCO impacted by high cost of proprietary hardware RAID solutions

G4: Platform I/O requirements, driven by memory size, outpace advancements made in storage technologies

G5: Deployed NFS file systems have long-standing scalability, performance, and access issues

Five-Year Plan

FY14

- Focus user support activities on lscratch1 in support of production CCCs on Sequoia
- Enhance ZFS-based Lustre metadata performance in support of user and purge performance
- Support the development, testing, and deployment of Lustre version 2.5 in classified and unclassified environments
- Migrate NetApp-based Lustre file systems to ZFS underpinnings
- Perform required development in support of ZFS software RAID solutions for Lustre object storage targets (OSTs)
- ~~Test and refine operational environment for JBOD-only Lustre file system deployments (G3)~~
- ~~Deploy initial pNFS NAS file systems (G4)~~
- ~~Continue HPC-focused development in concert with Lustre community (G1, G2, G3)~~
- ~~Field new file system(s) in support of ASC platform requirements (G3)~~

FY15

- Deploy ZFS software RAID-based Lustre file system into production (G3)
- Deploy initial pNFS NAS file systems (G5)
- Continue HPC-focused development in concert with Lustre community (G1, G2, G3)
- Field new file system(s) in support of ASC platform requirements (G3)
- ~~Support production implementation of tri-lab shared file system (G4)~~

FY16

- Continue HPC-focused development in concert with file system community (G1, G2, G3)
- Field new file system(s) in support of ASC platform requirements (G3)
- Prepare file system solution for Sierra ATS (G4)
- Support production implementation of tri-lab shared file system (G5)

FY17

- Continue HPC-focused development in concert with file system community (G1, G2, G3)
- Field new file system(s) in support of ASC platform requirements (G3)
- Deploy file system for Sierra ATS (G4)

FY18

- Continue HPC-focused development in concert with file system community (G1, G2, G4)
- Field new file system(s) in support of ASC platform requirements (G4)

Networking and Test Beds (LLNL)

The Networking and Test Beds project provides research, performance testing, capability testing, and analysis for the file system, network, and interconnect subsystems in support of current and future systems and environments. This work relies heavily on an adequately provisioned test bed, skilled staff, and collaborations with vendors.

This project will test various hardware and software components to quantify the features, performance, reliability, security, and interoperability of the products and broader technology base. The information acquired as a result of this project will be used to help determine an integrated architecture and resultant procurements for these subsystems.

Required Capabilities

R1: Adequately provisioned test bed for researchers and developers to perform testing and analysis in support of current and future systems and environments

Gaps

G1: Limited or lack of availability of next-generation CPUs and GPUs ~~technologies only recently on the market~~ (for example, ARM64, Haswell, Atlas, Kaveri) ~~Intel KNC and Nvidia Kepler)~~

G2: Lack of adequate Lustre/IB (Infiniband) monitoring tools

G3: Infiniband distance limitations

G4: Lack of IB congestion control

G5: File system Total Cost of Ownership (TCO) unsustainable into the future

Five-Year Plan

FY14

- Evaluate additional JBOD hardware for use with Lustre and develop tools to manage JBODs
- Evaluate Open Compute Platform for use as compute, data, and Lustre resource
- Evaluate ARM64 as compute platform, and collaborate with RedHat to address deficiencies in software
- Evaluate Aries interconnect on commodity hardware
- Perform RedHat Enterprise 7 alpha testing
- Develop software in support of next-generation resource manager
- Evaluate AMD products: CPUs with host serial adaptor (HAS), discrete GPUs
- Continue evaluating and testing Infiniband Federated data rate (FDR)
- Evaluate production release of Mellanox MetroX
- ~~Evaluate new storage hardware technologies~~

- ~~Perform compatibility testing with new IB hardware~~
- ~~Evaluate different network topologies for cluster interconnects (G3)~~

FY15

- Evaluate Infiniband EDR technology (G3)
- Evaluate new processor and GPU technology

FY16

- Evaluate new processor technology

FY17

- Evaluate new memory architectures, if available

FY18

- Evaluate new processor technology; evaluate creating HPC-specific CPU based on licensed IP

File Systems, Archival Storage, and Networking (LANL)

~~Archival and File Systems (LANL)~~

Capabilities of the Archival and File Systems components of the project ~~provides design and development of archival services, and high-performance file systems and I/O infrastructure for the ASC program. Capabilities supported~~ include online file systems such as the NFS complex and enterprise-wide supercomputer file systems, GPFS development, deployment and management, scalable I/O (SIO) middleware development and support, interconnect technology development and deployment, SAN development and deployment, and archive.

The file systems element of the project provides end-to-end, high-performance networking and SIO infrastructure for the ASC program. Successfully meeting the ASC programmatic milestones requires carefully balanced environments in which the I/O infrastructure scales proportionally with increased ASC platform capabilities and application data needs. As the program moves toward exascale areas, these efforts will improve the scaling or programmability of the I/O in ASC applications for current and future large-scale machines. Current areas of investigation are Parallel Logged File System (PLFS), scalable indexing, burst buffer architectures, and scalable metadata.

Application Readiness capabilities are consolidated in this project, addressing issues with an application's production-run readiness on current and incoming computing systems at LANL. Working with subsystem teams such as systems management; file systems; and I/O, archive, and tools, the Application Readiness team identifies causes of unexpected behavior and deploys fixes in production so that system users are able to make productive use of the systems with their applications to solve their problems. The team provides production problem solving (create small problem reproducers, identify cause, consult with the relevant technical experts to find a solution, and verify the deployed solution), periodic stress testing/regression of production machines, new software version regression testing, system configuration verification and software stack deployment with real user applications and metrics, and analysis/profiling.

The project also includes software support capabilities focused on communication and networking libraries (MPI). The goal is to establish a strong development and analysis tool capability for current and next-generation HPC platforms, including parallel capabilities. The project is focused on working with the HPC tool community and vendors to identify, plan, and integrate tools into production environments and establish a solid support structure.

Required Capabilities

R1: Provide design and development of archival services, high-performance file systems, and I/O infrastructure for ASC program

R2: Provide high-performance networking and SIO infrastructure for the ASC program

R3: Improve the scaling or programmability of I/O in ASC applications for current and future systems (PLFS, scalable indexing, burst buffer, archive, and scalable metadata)

R4: Ensure application readiness on all LANL and ACES systems, with workloads performing at high efficiency and reliability

~~R4: Provide archive system and tools for moving data to that system~~

R5: Strong MPI development and support, with strategic/support plans and collaborations to continue to increase scalability and support of programming environment and analysis tools

Gaps

G1: System memory growth exceeds growth of bandwidth to storage (Trinity will have ~5 PB and exascale systems will have ~32-64 PB)); checkpointing must be faster (30 PB over 5 minutes per hour at 100 TB/sec); scratch file systems must be much larger (approximately 30x system memory)

G2: Mean Time to Interrupt (MTTI) is getting smaller as clusters get larger (Trinity will have 10,000s of nodes and exacale systems will have 100,000s to 1,000,000s)

G3: Archives grow by about 3x system memories per month; the volume is handled through increases in density, but bandwidth to the archive is now an issue

G4: Applications need an I/O abstraction for the underlying file system

G5: Data about usage statistics, characteristics, and performance of systems and applications at scale are lacking

G6: HPC systems are becoming larger and more complex with a diversity of technologies and programming models, which stresses the scalability of applications and the capabilities of diagnostic tools and techniques

G7: Communication libraries and supporting software tool advances are required to maintain scalability and performance of rapidly changing programming models and HPC architectures

~~G3: Checkpointing must be faster (30 PB over 5 minutes per hour at 100 TB/sec)~~

~~G4: Scratch file systems must be much larger (approximately 30x system memory)~~

~~G5: Archives grow by about 3x system memories per month; the volume is handled through increases in density, but bandwidth to the archive is now an issue~~

~~G6: Applications need an I/O abstraction for the underlying file system~~

~~G7: N-N I/O at exascale will overload file system metadata servers~~

~~G8: Data about usage statistics, characteristics, and performance at scale~~

Five-Year Plan

FY14

- Provide on-going support and testing for production file-systems and HPSS
- Perform an assessment of network file system (NFS) with respect to limiting data movement
- Explore lower-cost open source or commercial archival solutions
- Provide application readiness support on capacity platforms
- Continue PLFS development under the EMC Corporation CRADA, including burst buffer integration and support for production
- Assist system management personnel with problem investigation and resolution
- Design and prototype a burst-buffer enablement library to insulate application developers from the complexities of next-generation file-systems
- Continue building an MPI support capability by engaging the community support model; focus will be on Open MPI development targeted to tri-lab needs, and interactions with threading models such as OpenMP
- Continue debugger and performance analysis support capability; focus will be on Open|SpeedShop and CBTF
- Initiate an OpenMP support capability
- ~~Perform and document a formal TCO analysis for disk-based archive (G1, G5)~~
- ~~Expand PLFS to other underlying file systems and/or models (G1, G6, G7)~~
- ~~Design modified to archival tools to support disk-based archive system (G1, G4, G5)~~
- ~~Perform and document a formal TCO analysis for a burst buffer (G1, G2, G3, G4)~~

FY15

- Perform initial hardware testing at scale using pre-exascale clusters (G1, ~~G2,~~ G3~~, G4, G6, G7~~)
- Test MTTI of new systems and measure actual performance characteristics of a production burst buffer (G2, G3~~, G8~~)
- Test N-1 and N-N tools for helping applications make use of burst buffer; evaluate advantages and disadvantages of both using actual hardware (G1, G5, G6, G8)
- Construct new archival prototype using FY13–FY14 work, test performance characteristics and scaling, and determine desired modifications for Trinity-sized long-term workloads (G1, G3, G4, G5)
- Support users of LANL and ACES systems by tackling hard-to-diagnose problems, typically involving the interaction of applications with multiple aspects of the computational environment (G6)
- Provide MPI development and user support and sustain capabilities for scalable programming environment tools for current and future systems (G7)

FY16

- Deploy I/O abstraction software for general use with the burst buffer (G1, G4~~, G5~~)
- Deploy a large-scale implementation of a future archive platform (G1, G3, ~~G4, G5~~)
- Collect public usage statistics for archive, burst buffer, and scratch file system (G5, ~~G8~~)
- Support users of LANL and ACES systems by tackling hard-to-diagnose problems, typically involving the interaction of applications with multiple aspects of the computational environment (G6)
- Provide MPI development and user support and sustain capabilities for scalable programming environment tools for current and future systems (G7)

FY17

- Refine I/O software based on usage statistics and lessons learned (G1, G4, G5)
- Publish collected statistics and participate in DOE (G5, ~~G8~~)
- Evaluate parallel file system performance for exascale machines (G1, G2, ~~G3,~~ G4, G5, ~~G7~~)
- Support users of LANL and ACES systems by tackling hard-to-diagnose problems, typically involving the interaction of applications with multiple aspects of the computational environment (G6)
- Provide communication library development and user support and sustain capabilities for scalable programming environment tools for current and future systems (G7)

FY18

- Plan and design parallel file system and archival storage strategy for exascale systems (G1, G2, G3, G4)
- Support users of LANL and ACES systems by tackling hard-to-diagnose problems, typically involving the interaction of applications with multiple aspects of the computational environment (G6)
- Provide communication library development and user support and sustain capabilities for scalable programming environment tools for current and future systems (G7)

Archival Storage (SNL)

The Archival Storage project represents SNL's participation in the DOE HPSS Consortium development project. HPSS provides the archival storage solution for ASC systems and is in direct alignment with ACES.

SNL's role in the HPSS project is to collaborate with tri-lab developers to design, implement, and test solutions that meet ASC requirements for all three labs.

Required Capabilities

R1: Current release levels of HPSS are stable and support existing capacity computing resources at Sandia

R2: HPSS performance enhancements to demonstrate sufficient file create rates and data ingestion speeds which will be needed to support future Advanced Technology Systems; these are estimated to be on the order of 30,000 file creates per second, and ingest rates of a few TB/s.

R3: Data integrity to ensure no loss of data

Gaps

G1: The current HPSS design is not sufficiently parallel to accomplish the performance gains required in the next decade; a new, more parallel architecture is planned for Release 8.x and beyond

G2: The Redundant Array of (Inexpensive) Tape (RAIT) feature being deployed at NCSA in HPSS 7.4 will improve the recoverability of data in the face of tape damage

G3: End-to-end data error checking will ensure accuracy of data transfers both to and from the HPSS archive systems(this feature is not yet in the design/deploy schedule)

Five-Year Plan

FY14

- Deploy HPSS 7.4 releases integrating RAIT and performance enhancements
- Continue development and testing for next-generation environments; continue to develop and tune partitioned database features with a target of 5000 file creates per second on a single system
- Negotiate ownership and support of the HSI/HTAR tools from Gleicher Enterprises for use by ASC
- ~~Deploy v7.p (G1)~~
- ~~Design and prototype v8.1 with partitioned database across multiple systems, with a target of 30,000 file creates per second (G1)~~

FY15

- Design and prototype v8.1 with partitioned database across multiple systems, with a target of 30,000 file creates per second (G1)
- Proceed with development and testing of v8.1 (G1, G2, G3)

FY16

- Deploy v8.1 (G1, G2, G3)
- Design and prototype v8.2 with core services distributed across multiple systems, with a target of satisfying all capabilities with ~100,000 file creates per second (G1, G2, G3)

FY17

- Proceed with development and testing of v8.2 (G1, G2, G3)

FY18

- Deploy v8.2 (G1, G2, G3)

Scalable Input/Output Research (SNL)

The SIO project provides support for I/O library and file systems on existing petascale ASC platforms as well as critical R&D to provide I/O capabilities on future exascale platforms. The research performed in this project directly addresses two vital concerns for I/O on exascale platforms: scalable parallel file systems and technologies for integration of computation and analysis.

To improve file-system scalability and resilience, the SIO project is developing a file system that decentralizes management of devices to support a high degree of heterogeneity within a system of inherently unreliable networks and storage devices. The central components of this peer-to-peer-like system are "smart" servers that have access to a variety of different local and remote media (for example, disk, NVRAM, memory, and tape) and are pervasive throughout the computing platform. These servers directly handle I/O requests, initiate third party transfers, or replicate the data as needed.

The SIO project is also developing technologies for integration of computation and analysis. This software will have tremendous impact on I/O for petascale and future systems because it allows for the creation of integrated scientific workflows that process data before the data reaches persistent storage, avoiding storage of transient data products. Current use of this technology includes data staging/caching to manage bursty I/O operations (for example, for checkpoints) and in-transit fragment detection for CTH, a co-design activity with the Scalable Data Analysis project.

Required Capabilities

R1: Scalable, fault-tolerant, parallel file system

R2: System software and libraries to support efficient data management and coupling of simulation and analysis

R3: Standard interfaces and programming models to leverage advanced architectures for data-intensive applications

Gaps

G1: Methods to reduce, eliminate, or hide the cost of data movement to external storage. Storage devices and file systems are projected to be a critical bottleneck in HPC processing for exascale. Current production environments lack the tools, programming models, and system software needed to address this issue. R&D in new approaches that exploit data locality, integration of simulation and analysis, I/O scheduling, and caching are needed.

G2: Resilience in extreme-scale storage systems and I/O middleware. The protection of application data has always been an extremely important requirement for I/O systems. In exascale environments where application data will likely be managed by large numbers of disparate resources with abundant device failures, we need to properly define semantics and expectations for data coherency and consistency.

G3: Expert knowledge of exascale storage environments. The tools needed to investigate and evaluate exascale storage approaches do not exist. There is a need for accurate simulation models to support failure cases and issues that only exist at large scales.

Five-Year Plan

FY14

- Develop and demonstrate a globally accessible data service for in-memory data storage (G1, G2)
- Develop extended feature support for the Sirocco file system, including storage servers that use gossip protocols to exchange information and log-based, on-disk storage (G1, G2)
- Analyze topological placement issues related to coupled codes and in-transit data services (G1)
- ~~Investigate data-migration verses task migration for data services; addresses critical data-locality issues expected to exist for exascale (G1)~~
- Explore resilience/durability of in-transit workflows, which ~~Integrate transaction-based methods into production systems (this~~ coincides with the completion of an early career LDRD on distributed transactions (G2)
- ~~Explore the use of publish/subscribe protocols for data services (G1)~~

FY15

- Modify scheduler placement algorithms to better utilize network usage for coupled applications (G1)
- Integrate in-memory data storage capabilities into other ASC-developed capabilities for programming models and resilience (G1)
- Develop dynamic selection and placement of data services; leverage graph-partitioning algorithms to address data locality (G1)

FY16

- Transition Sirocco file system into production (G1, G2)

FY17

- Deploy data services with system support for placement in production environment (G1)

FY18

- Adapt data services to fully support code coupling, integrated analysis, and resilience in a prototype exascale environment (G3)

Scalable Interconnects for Extreme-Scale Tightly Coupled Systems (SNL)

The Scalable Interconnects for Extreme-Scale Tightly-Coupled Systems project will develop capabilities to enable performance and scalability of ASC applications on current and future high-performance interconnection networks on extreme-scale platforms. This project will concentrate on characterizing application requirements with respect to functionality and performance for intra-application data movement as well as application network transfers to external I/O services. It will also provide a low-level network programming interface appropriate for current-generation network hardware as well as more advanced next-generation hardware with more sophisticated network interface capabilities and functionality. As applications explore alternative programming models beyond the current distributed memory MPI model, the low-level network programming interface must evolve to include the ability to provide very lightweight one-sided data transfer operations, while continuing to enable efficient two-sided message-based transfers. It is likely that this project will expand to include an analysis of network topologies, network interface hardware design and evaluation, optimized network transfer protocols, and system software support for advanced network interface operations.

This project will build on existing efforts surrounding the development of the next-generation portals network programming interface and measurements of application sensitivity to network performance. It has close ties to the ACES project and the Cray network hardware design and engineering activity in WBS 1.5.4.1.

Required Capabilities

R1: Future interconnects for extreme-scale platforms must enable performance and scalability for ASC applications

R2: A low-level interconnect programming interface that meets the functionality, performance, and scalability requirements of applications and system services

Gaps

G1: There is currently very little data that describes or characterizes the communication requirements of ASC applications in a way that allows for making informed decisions about interconnect hardware and software design and implementation.

G2: Many-core processors and other factors such as energy and resilience are motivating explorations of alternative parallel programming models beyond MPI, and the interconnect hardware and software needs to be able to support a broader spectrum of functionality. System software will also require more advanced features from the interconnect to be able to provide the services and support necessary for these alternative parallel programming models.

Five-Year Plan

FY14

- Complete a study to determine the appropriate ratio of PPE threads to MPI ranks for the Intel Phi platform (G1, G2)
- Implement and evaluate the effectiveness of using SIMD vectorization and other node-level hardware capabilities to accelerate MPI tag matching (G1, G2)
- Release an enhanced version of the Portals 4.0 reference implementation (G2)
- ~~Use existing large-scale platforms to characterize the interconnect bandwidth requirements of scalable applications employing mixed parallel programming models (G1)~~

FY15

- Continue to develop and enhance portals to meet the ongoing interconnect requirements of applications and services (G2)

FY16

- Provide a set of interconnect performance requirements for important ASC applications (G1)

FY17

- Demonstrate a hardware implementation of portals (G2)

FY18

- Complete a study that evaluates the performance of the hardware implementation of Portals (G1, G2)

Post-Processing Environments (WBS 1.5.4.6)

This level 4 product provides integrated post-processing environments to support end-user visualization, data analysis, and data management. The scope of this product includes planning, research, development, integration and deployment, continuing customer/product support, and quality and reliability activities, as well as industrial and academic collaborations. Projects and technologies include tools for metadata and scientific data management, as well as general-purpose and application-specific visualization, analysis, and comparison. Research includes innovative data access methods and visualization of massive, complex data—the use of open-source foundations will continue to be an important strategy for development of shareable advanced techniques. The product must develop solutions to address interactivity, scaling, tri-lab access for petascale platforms, and data analysis techniques needed to support effective V&V and comparative analysis. Solutions for emerging platform architectures may in turn require customization and/or re-architecting of software to leverage hardware features. A continuing emphasis will be placed on tools for improving end-user productivity. The product also provides and supports infrastructure including office and collaborative space visualization displays, mechanisms for image data delivery, and graphics rendering hardware.

WBS 1.5.4.6 Scientific Visualization (LLNL)

The Scientific Visualization project conducts research and develops and supports tools for managing, visualizing, analyzing, and presenting scientific data. Research topics include topological analysis, particle visualization, and data compression techniques. Operational support for data analysis covers support of post-processing resources, including visualization servers, displays, and facilities. The visualization hardware architecture team engages in planning, test bed prototyping, testing of systems and components, and procurement and integration of new systems. Display efforts include support of high-resolution, high-performance display devices for theaters and collaborative use areas. The project installs, maintains, and consults on software visualization tools, and supports demonstrations on the PowerWalls. The project maintains unclassified and classified video production labs and consults on software such as resource management tools, movie players, animation, and visualization packages. The project exploits the latest capabilities of clustering hardware, GPUs, and parallel storage systems. Hardware capabilities include three production visualization servers and several PowerWall clusters. A video display infrastructure drives PowerWalls and smaller displays. Visualization researchers continued to perform work in areas of topology, compression, and advanced data analysis techniques.

Required Capabilities

R1: Data analysis hardware and software to support the analysis and post-processing of ASC simulation data

R2: PowerWalls for presenting high-resolution animations and images of large-scale data and operational support for high-level demonstrations

Gaps

G1: GPU advances are needed to further scale down the platforms driving PowerWalls and to enable more efficient data analysis

G2: Data analysis and visualization techniques to successfully analyze data generated on post-petascale platforms

Five-Year Plan

FY14

- Provide consulting and maintenance for the data analysis and visualization hardware platforms and software environment
- ~~Enable the use of GPU programming on Edge through support of CUDA, OpenCL, and other acceleration programming capabilities and training (G1)~~
- Provide initial start up and ongoing ~~Continue to maintain the~~ data analysis efforts, including the creation of movies and visuals ~~and visualization hardware platforms and software environment~~

- Provide operational support for ~~all visualization facilities, including supporting~~ projection systems and other equipment associated with ASC visualization theaters ~~and facilitating the use of the data analysis clusters and associated storage~~
- Support large-scale visualization and data analysis efforts ~~activities~~, including ~~supporting ASC scientists with~~ the creation of movies and visuals ~~for presenting and analysis scientific data~~
- Develop an initial flow visualization capability using steam-line tracing algorithms to be deployed in Lorenz using WebGL
- Prepare for next-generation data analysis through research in multi-resolution techniques, data compression, and topological methods
- ~~Exploit research results in data analysis and visualization to prepare for post-petascale analysis (G2)~~

FY15

- Evaluate, select, and deploy visualization hardware to support ASC data analysis needs
- Continue to maintain the data analysis and visualization hardware platforms and software environment and provide operational support for all visualization facilities, including supporting projection equipment and facilitating the use of the data analysis clusters and associated storage.
- Support large-scale data analysis and visualization activities, including supporting ASC scientists with creation of visuals and movies for presenting and analyzing scientific data
- Exploit research results in data analysis and visualization for ASC simulations

FY16

- Maintain the data analysis and visualization hardware platforms and software environment and provide operational support for visualization facilities
- Refresh PowerWall driver technology as needed

FY17

- Maintain the data analysis and visualization hardware platforms and software environment and provide operational support for visualization facilities
- Support large-scale data analysis and visualization activities

FY18

- Maintain the data analysis and visualization hardware platforms and software environment and provide operational support for visualization facilities
- Support large-scale data analysis and visualization activities

- Gather requirements, procure, and support data analysis and visualization platform to support the 2017 Sierra ATS

Scientific Workflow and Data Management (LLNL)

The Scientific Workflow and Data Management project provides users with powerful and time-conserving ways to access, search, compare, and archive large-scale scientific data, and new high-level tools for managing the simulation workflow. This is achieved through the development of production-quality applications that enhance data management capabilities and the creation of innovative interfaces to job monitoring and vertical application frameworks.

Hopper and Chopper are the principal products of the data management effort. In the simulation workflow area, the Lorenz Web-based HPC application suite forms a foundation for providing new ASC-specific capabilities. Lorenz uses advanced Web technologies to make HPC more accessible, saving the user time while also helping the resources to be used more effectively.

Required Capabilities

R1: Flexible data management software that helps users make efficient and appropriate use of resources

R2: Highly customizable and easily accessible user interface to HPC resources and information

~~R3: Robust tools for supporting the creating and sharing of complex workflows~~

Gaps

G1: Preparation for exascale data management limited until preferred exascale data storage techniques emerge

G2: Using a computer center effectively still requires significant understanding of local policies, restrictions, and implementations

G3: Computer center information, ranging from hardware/software status to policies, is still disjoint

G4: Workflow management expertise required from collaborators from other projects

Five-Year Plan

FY14

- Release new versions of Hopper and Chopper with a focus on helping users deal with exceptionally large collections of data, including file transfer wizard that optimizes the transfer path in terms of both speed and efficiency of resources
- Investigate the inclusion of RobinHood-based Lustre metadata in Hopper and Chopper, with the expectation of significantly improving directory list and disk usage operations in Lustre file systems
- Develop and release new versions of Lorenz that support the growing number of users who are turning to Web-based tools for interacting with the computing center

- Extend the MyLC dashboard to include more complete information about changes that are occurring within the center, and provide users with a variety of ways to subscribe to this information
- ~~Consolidate and integrate computer center data collection, distribution, and display into a common framework (G3)~~
- ~~Incorporate workflow management capabilities, leveraging existing work and extending where needed (G4)~~

FY15

- Incorporate hardware monitoring information into Lorenz, giving users a real-time view of the status of network, file system, and computing systems (G3)
- ~~Explore new protocols and techniques for managing extreme scale data management (G1)~~

FY16

- Explore new protocols and techniques for managing extreme scale data management (G1)
- ~~Implement and coordinate workflows that are tuned to exascale, allowing users to leverage the work of domain experts (G4)~~

FY17

- Incorporate support for the dominant exascale data storage and handling protocols into high-level tools for end users (G1)

FY18

- Implement and coordinate workflows that are tuned to exascale, allowing users to leverage the work of domain experts (G4)

Visualization and Data Analysis (LANL)

~~Production Visualization (LANL)~~

Data analysis and visualization are key capabilities in taming and understanding the ever-increasingly large datasets generated from extreme-scale scientific simulations. This project comprises research, development, deployment of software and facilities to production and ongoing expert support in this.

The production and facilities component of the ~~purpose of the Production Visualization~~ project is to provide LANL weapons designers with visualization systems research and support, and to provide analytic expertise to help LANL weapons designers utilize the full power of the hardware and software infrastructure for visualization and data analysis developed and deployed by ASC, thus improving the physics understanding of their weapons simulations.

Technical staff members funded by the project ~~Production Visualization~~ assist in the design theater and the co-laboratories. The project also deploys within the design community in X Division a small group of individuals with expert knowledge in both visualization and weapons science to work directly with the designers. ~~The project's~~ Capabilities include the design and deployment of new visualization systems, briefing support, and support of large facilities, such as the CAVE and the PowerWall. The project provides assistance, training, and developing of new tools to work with these facilities. Development, deployment, and maintenance of any needed visualization corridor software are also provided by this project.

The project is responsible for the EnSight visualization and data analysis software, including maintaining the EnSight software installation laboratory-wide, providing local user support in the use of the software, and acting as a bridge between the LANL design community and the EnSight developers at Computational Engineering International.

The R&D aspect of the project develops new visualization algorithms and systems to meet current and future capability requirements for ASC simulations. This work is required to address ASC workloads: massive data sizes, complex results, and the use of unique supercomputing architectures.

ASC simulations are currently producing massive amounts of data that threaten to outstrip the ability to visualize and analyze it. Therefore, it is important to understand how to triage data within the simulation as it is generated using techniques such as *in-situ* analysis for data reduction and visualization kernels that run on the supercomputing platform, including data analysis, visualization, and rendering methods. Building on the Level 2 milestone work in FY13 to demonstrate *in situ* techniques in IC, the project will extend this capability to other codes and to enabling the design community to explore the new level of analytic flexibility now available.

Another important analysis mode involves storing reduced data for later interactive analysis and visualization (that is, post-processing analysis). Extreme-scale ASC databases, petabytes or larger, already exist and are growing in number and size. LANL's long-term objective is to build a post-processing analysis system that, within reasonable limits, can manage performance optimization tasks automatically, be fault-tolerant, provide a set of high-level serial and parallel programming primitives for carrying out complex queries and computations on the stored data, and provide fast enough execution to enable interactive, iterative discovery. Current approaches range from the database-driven (SQL) to parallel computation-driven (Hadoop/MapReduce), neither of which alone can meet the needs of the target community. LANL will build on these technologies, integrating the best of both approaches.

Required Capabilities

R1: Visualization computing systems and facilities operating at high availability

R2: Support for visualization computing in the advanced and production computing systems, and expert visualization support for ASC simulation scientists

R3: *In-situ* integration with scientific simulation codes

R4: Quantifiable data reduction and triage

R5: Portable, scalable visualization and analysis software and hardware infrastructure

~~R3: Visualization facilities operating at high availability~~

~~R4: Expert visualization support for ASC simulation scientists~~

Gaps

G1: Inadequate post-processing capabilities must be developed for the exascale regime and for new GPU technologies

G2: Lack of bandwidth to storage

G3: Lack of deep, quantitative understanding of massive simulation data

G4: Lack of portable, efficient, scalable visualization and analysis software infrastructure for emerging next-generation architectures

G5: Lack of visualization and analysis software for ensembles, coupled multiscale, and multivariate physics

~~G2: Reads from storage must be faster, to accommodate larger data~~

~~G3: Changes in GPU technology must be exploited~~

~~G4: Burst buffer use must be explored and exploited~~

Five-Year Plan

FY14

- Continue to support and maintain production visualization systems
- Provide contract management and requirements specification, including facilities, visualization cluster, and EnSight contracts

- Provide technical guidance on visualization needs for the forthcoming Trinity ATS
- Work directly with designers in physics-based, iterative discovery process using EnSight to promote new discoveries in weapons science in programs
- Extend PISTON, a portable hardware-accelerated visualization library, to Blue Gene, GPU, and Intel MIC, including distributed memory and unstructured mesh support (work with Kitware for integration into open source tools)
- Apply *in situ* data analysis framework to additional ASC codes
- Evaluate data-intensive computing technology for ASC programmatic needs
- Evaluate optimization strategies for efficient data flow in visualization, analysis, and parallel storage systems
- Provide additional expert visualization support to designers and analysts
- ~~Continue to promote new discoveries in weapons science by advanced applications of visualization and data analysis~~
- ~~Continue to document the results with classified papers and publications on weapons science topics jointly co-authored with X Division designers~~

FY15

- Develop visualization infrastructure for Trinity (G1)
- Continue to support and maintain production visualization systems and to promote new discoveries in weapons science by advanced applications of visualization and data analysis (G1, G2, G3)
- Develop data triage and reduction capability for large-scale visualization and analysis with automatic quantifiable metrics on the data output (G2)
- Develop a visualization and analysis capability that generates quantified analysis products from large-scale inputs via quantified data triage and reduction (G2)
- Deploy a data-intensive analytics solution that integrates with ASC simulation codes, occurring in conjunction with participation in ASC Trinity system procurement and roll-out (G1, G3)
- ~~Deploy Trinity~~
- ~~Plan and design new visualization compute system, if needed~~
- ~~Continue to support and maintain production visualization systems~~
- ~~Continue to promote new discoveries in weapons science by advanced applications of visualization and data analysis~~

FY16

- Continue to support and maintain production visualization systems and to promote new discoveries in weapons science by advanced applications of visualization and data analysis (G1, G2, G3)
- Extend PISTON to include a sufficiently comprehensive suite of commonly used visualization and analysis operators and a sufficiently accessible user interface such that it can be used by domain experts in a wide array of scientific applications (G1, G3)
- Explore, develop, and deploy quantitative visualization and analysis for ensembles, coupled multiscale, and multivariate physics (G4)
- ~~Deploy new visualization compute system, if needed~~
- ~~Continue to support and maintain production visualization systems~~
- ~~Continue to promote new discoveries in weapons science by advanced applications of visualization and data analysis~~

FY17

- Continue to support and maintain production visualization systems and to promote new discoveries in weapons science by advanced applications of visualization and data analysis (G1, G2, G3)
- Deliver an integrated software infrastructure for visualization and analysis applied to ASC codes of interest as part of a Level 2 milestone (G1, G2, G3, G4)
- ~~Test exascale technologies~~
- ~~Continue to support and maintain production visualization systems~~
- ~~Continue to promote new discoveries in weapons science by advanced applications of visualization and data analysis~~

FY18

- Continue to support and maintain production visualization systems and to promote new discoveries in weapons science by advanced applications of visualization and data analysis (G1, G2, G3)
- Explore data analysis and visualization technologies for and integrated hardware and software infrastructure for extreme-scale systems (G1, G2, G3, G4)

~~*Visualization and Data Analysis Research and Development Project (LANL)*~~

~~The Visualization and Data Analysis R&D Project develops new visualization algorithms and systems to meet capability requirements for ASC simulations. This work is required to address ASC workloads: massive data sizes, complex results, and the use of unique supercomputing architectures.~~

~~ASC simulations are currently producing massive amounts of data that threaten to outstrip the ability to visualize and analyze it. Therefore, it is important to understand how to triage data within the simulation as it is generated using techniques such as *in-situ* analysis for data reduction and visualization kernels that run on the supercomputing platform, including data analysis, visualization, and rendering methods.~~

~~Another important analysis mode involves storing reduced data for later interactive analysis and visualization (that is, post-processing analysis). Extreme-scale ASC databases, petabytes or larger, already exist and are growing in number and size. LANL's long-term objective is to build a post-processing analysis system that, within reasonable limits, can manage performance optimization tasks automatically, be fault-tolerant, provide a set of high-level serial and parallel programming primitives for carrying out complex queries and computations on the stored data, and provide fast enough execution to enable interactive, iterative discovery. Current approaches range from the database-driven (SQL), to parallel computation-driven (Hadoop/MapReduce), neither of which alone can meet the needs of the target community. LANL will build on these technologies, integrating the best of both approaches.~~

~~*Required Capabilities*~~

~~R1: In-situ integration with scientific simulation codes~~

~~R2: Quantifiable data reduction and triage~~

~~R3: Portable, scalable visualization and analysis software and hardware infrastructure~~

~~*Gaps*~~

~~G1: Lack of bandwidth to storage~~

~~G2: Lack of deep, quantitative understanding of massive simulation data~~

~~G3: Lack of portable, efficient, scalable visualization and analysis software infrastructure for emerging next generation architectures~~

~~G4: Lack of visualization and analysis software for ensembles, coupled multiscale, and multivariate physics~~

~~*Five-Year Plan*~~

~~FY14~~

- ~~Develop data triage and reduction capability for large-scale visualization and analysis with automatic quantifiable metrics on the data output (G2)~~

- ~~Develop a visualization and analysis capability that generates quantified analysis products from large-scale inputs via quantified data triage and reduction (G2)~~

~~FY15~~

- ~~Deploy a data-intensive analytics solution that integrates with ASC simulation codes, occurring in conjunction with participation in ASC Trinity system procurement and roll-out (G1, G3)~~

~~FY16~~

- ~~Extend PISTON to include a sufficiently comprehensive suite of commonly used visualization and analysis operators and a sufficiently accessible user interface such that it can be used by domain experts in a wide array of scientific applications (G1, G3)~~
- ~~Explore, develop, and deploy quantitative visualization and analysis for ensembles, coupled multiscale, and multivariate physics (G4)~~

~~FY17~~

- ~~Deliver an integrated software infrastructure for visualization and analysis applied to ASC codes of interest as part of a Level 2 milestone (G1, G2, G3, G4)~~

Scalable Data Analysis (SNL)

The Scalable Data Analysis project provides data analysis capabilities and support for a range of SNL ASC customers—from analysts and code developers to algorithm designers and hardware architects. Capabilities include data manipulation, data transformation, and data visualization that contribute to insight from computational simulation results, experimental data, and/or other applicable data. A project emphasis is to deliver and support scalable capabilities that support increasing data sizes, data sources, and platform processor counts for ASC complex applications and system architecture.

This project includes production deployment and support services that enable ASC customers to carry out data analysis on ASC systems. This includes porting and installation of tools onto production systems; maintenance, testing, debugging, refinement and integration of tools in the end-to-end system environment as needed to assure effective end-user capabilities; and user support. SNL priorities include a focus on delivering and supporting analysis capability for Cielo and subsequent ACES platforms.

Current tools include scalable data analysis software released open source through ParaView and the VTK, an early-release *in-situ* data analysis library (Catalyst) for coupling directly with running simulation codes, and R&D prototypes for the analysis of results from ensembles of simulation runs. Current hardware platforms for data analysis are limited to data analysis/visualization partitions on the compute platforms with an emphasis on delivery of visualizations to desktop.

Partnering with ASC customers and other product areas, this project will continue to build on its successful ParaView and VTK-based products. The project performs R&D that advances these capabilities as needed for evolving next-generation architectures, ensuring that ASC's investment in data analysis and visualization will provide advanced capabilities on platforms from Cielo through future exascale systems.

Required Capabilities

R1: Scalable post-processing tools for analysis and visualization on existing and next-generation architectures

R2: Scalable *in-situ* data analysis and visualization capability to mitigate impact of I/O mismatch and power constraints at extreme scales

R3: Dedicated hardware resources for interactive data analysis and visualization

R4: Production deployment and support for interactive applications, *in-situ* analysis capabilities, and infrastructure

Gaps

G1: Algorithms, libraries, and tools that leverage future architectures

G2: Production-ready scalable *in-situ* data analysis and visualization capability, integrated into ASC codes, with production support, including quantifiable data reduction techniques and co-designed analysis capabilities for critical ASC decision support

G3: Ensemble and high-dimensional data analysis and visualization capabilities for extreme scale, heterogeneous output from *in-situ* analysis

Five-Year Plan

FY14

- Deliver initial ensemble analysis on the classified network for use in sensitivity analysis for electrical circuit simulations (G3)
- Develop and integrate select many-core algorithms into SNL production toolset (G1)
- Develop an *in-situ* capability using Catalyst for applications using the SIERRA toolkit (G2)
- ~~Develop multi-mode *in-situ* processing (both tightly coupled and loosely coupled) through common API; perform experimentation into scheduling, power, and productivity impacts (G2)~~
- ~~Integrate multi/many core algorithms with production toolset (G1)~~
- Continue *ParaView* and *Catalyst* releases, with production support in conjunction with Kitware, Inc. (G1)
- Deliver scalable analysis and visualization capabilities for Cielo and Sequoia customers (G1)

FY15

- Develop production tools to support Trinity
- Perform experiments on optimization of workflows with *in-situ* analysis capabilities on burst-buffer architectures (G2)
- Prototype integration of multi/many core code into production tools (G1)
- Continue *ParaView* and *Catalyst* releases, with production support (G1)

FY16

- Deliver analysis and visualization tools on Trinity for Limited Availability (G1)
- Perform experiments of multi-mode *in-situ* processing at scale on Trinity (G2)

FY17

- Continue *ParaView* and *Catalyst* releases, fully transitioned to multi/many core (G1)
- Release pre-production versions of Web-based post processing tools for ensemble and *in-situ* analysis results (G2)

FY18

- Release production versions of post-processing tools for ensemble and *in-situ* analysis results (G2, G3)

Facility Operations and User Support (WBS 1.5.5)

This sub-program provides both necessary physical facility and operational support for reliable, cross-lab production computing and storage environments as well as a suite of user services for effective use of ASC tri-lab computing resources. The scope of the facility operations includes planning, integration and deployment, continuing product support, software license and maintenance fees, procurement of operational equipment and media, quality and reliability activities, and collaborations. FOUS also covers physical space, power and other utility infrastructure, and LAN/wide area network (WAN) networking for local and remote access, as well as requisite system administration, cyber-security, and operations services for ongoing support and addressing system problems. Industrial and academic collaborations are an important part of this sub-program.

User Support Services (WBS 1.5.5.2)

This level 4 product provides users with a suite of services enabling effective use of ASC tri-lab computing resources. The scope of this product includes planning, development, integration and deployment, continuing product support, and quality and reliability activities collaborations. Projects and technologies include computer center hotline and help-desk services, account management, Web-based system documentation, system status information tools, user training, trouble-ticketing systems, and application analyst support.

Hotlines and System Support (LLNL)

The Hotlines and System Support project provides users with a suite of services enabling effective use of ASC computing resources for the tri-lab as well as academic and industrial collaborations. This project includes computer center hotline and help desk services, account management, Web-based system documentation, system status information tools, user training, incident management systems, and application analyst support. Services are provided to both LLNL users as well as users from external sites, including LANL, SNL, and the ASC Alliance sites.

This project provides accounts administration, technical consulting, and documentation and training to facilitate the effective use of LLNL HPC systems. An accounts specialist team provides all account management services necessary for users to obtain accounts and access LLNL HPC systems. This includes account creation and removal, bank allocations, token management and visitor tracking for foreign national users. The technical consultant team provides technical support to LLNL users to enable their effective use of LLNL HPC systems. Consulting services vary from helping new users configure their environment, assisting experienced users with optimization of codes, and supporting other Livermore Computing (LC) staff with monitoring of file systems, batch queues, and user environments. Extensive Web documentation, user manuals, technical bulletins, and training are provided to users via email, Web, and in-person training.

Required Capabilities

R1: Provide user account management, bank allocations, and One Time Password (OTP) support to all users of LLNL supercomputers

R2: Provide technical consulting services to all users of LLNL supercomputers

R3: Provide current technical documentation and training materials on the use of LLNL supercomputers

R4: Create and maintain tools for LLNL users that facilitate the effective use LLNL supercomputers

Gaps

G1: Succession planning is needed to respond to potential retirements of senior technical consultants

G2: Succession planning is needed to respond to the migration from the LC Identity Management system to a new Identity Management system

G3: Account processes are not fully automated for every special case (Remote, FN, SFN)

G4: A continually changing environment needs continuous updates in documentation, correctness, and status for excellent communication with user community

G5. New systems require user assistance in areas such as code porting and compiler nuances

Five-Year Plan

FY14

- Continue to provide ongoing support services for hotline operations, documentation, and training
- Continue to migrate ~~Convert~~ existing Livermore Computing (LC) Websites to ~~Web pages into~~ the new LLNL standard Website format ~~Web infrastructure~~ (G4)
- Provide support to the new DoD SIPRNet (SNSI) user community
- Deploy new whiteboard, system status, and bank allocation tools for improved user communication and hotline operations
- Continue to assist users in the migration of applications to the BlueGene/Q architecture
- Deploy a secure computing facility instance of ~~Prepare for the migration from~~ the Identity Management System ~~to a new tool~~ for managing user accounts, groups, and bank memberships ~~and banks~~ (G2)
- ~~Develop training and documentation for new TLCC3 systems (G4)~~
- ~~Automate existing manual account procedures (G3)~~

FY15

- Assist users in the migration of applications to the new CTS-1 ~~TLCC3~~ systems (G5)
- Migrate from the LC Identity Management system to a new tool for managing user accounts and banks (G2)
- Develop a new knowledge base of information for LC staff and users (G4)

FY16

- Develop training and documentation for the new ATS-2 system (G4)
- Deploy the new knowledge base of information for LC staff and users (G4)

FY17

- Assist users in the migration of applications to the ATS-2 system (G5)

FY18

- Continue to provide ongoing support services for hotline operations, documentation, and training (G1, G2, G4)
- Assist users in the migration of applications to the ATS-2 system (G5)

Integrated Computing Network Consulting, Training, Documentation, and External Computing Support (LANL)

The Integrated Computing Network Consulting, Training, Documentation, and External Computing Support project is responsible for direct customer service for local and remote users of ASC/LANL resources, the development and delivery of documentation and training materials for ASC/LANL resources, usage statistics, and an administrative interface for the ASC tri-lab, Alliance users, and other external ASC/HPC users. The primary capabilities consist of user support services, operational metrics for an HPC environment on, for example, usage and availability, Web-page development to present this information to system personnel and users, and the development of user documentation and training.

Required Capabilities

R1: Ongoing user support services

R2: Operational metrics for HPC resource usage and availability

R3: Development of user documentation and training

R4: Web-page development providing access to HPC services

Gaps

G1: Lack of experienced personnel resources and tools for continual support of expanding user needs for ASC/LANL/ACES computing resources

G2: Lack of details on evolving architectures and associated software environments to assist with Web-based system documentation development

Five-Year Plan

FY14

- Continue to enhance user support tools, capabilities, and infrastructure available to users and the user support team (G1)
- Develop Trinity (ATS-1 system) system documentation and training materials (G2)
- Provide assistance for use of emerging ASC platforms/architectures (G2)

FY15

- Develop CTS-1 ~~TLCC3~~ system documentation and training materials (G2)
- Provide ongoing support services, documentation, and training for ASC platforms/architectures

FY16

- ~~Develop exascale system documentation and training materials (G2)~~
- Provide assistance for use of emerging ASC platforms/architectures

FY17

- Continue to enhance user support tools, capabilities, and infrastructure available to users and the user support team
- Provide assistance for use of emerging ASC platforms/architectures

FY18

- Develop next-generation system documentation and training materials (G2)
- Continue to enhance user support tools, capabilities, and infrastructure available to users and the user support team

User Support (SNL)

The User Support project provides user support and associated resources for SNL computing systems and tri-lab resources. User support activities focus on improving the productivity of the entire user community, local and remote, in utilizing the ASC HPC resources.

This project deploys and maintains the following SNL capabilities for user support: 1) coordination between user support activities and leadership in adopting ITIL principles and practices; 2) ITIL-compliant incident, problem, and knowledge management tool set; 3) training facilities and equipment; and 4) a Web portal for HPC-related information, real-time data, and documentation.

In addition, this project provides the following user support capabilities in conjunction with other projects: 1) a tiered user support structure (HPC service desk) that responds to SNL and tri-lab user requests received via phone, email, Web-based requests, and in-person visits; 2) the Synchronized Account Request Automated Process (SARAPE) tri-lab account provisioning Web-based tool; 3) Web-based, classroom, and one-on-one training; and 4) direct support in utilizing ASC resources.

This project also funds the SNL user support team's involvement in collaborative efforts such as the Predictive Science Academic Alliance Program (PSAAP) and ACES.

Required Capabilities

R1: HPC service desk resources and infrastructure

R2: Experienced technical staff and support personnel versed in various aspects of HPC, including hardware and software environments, effective application facilitation, and use of all ASC HPC resources

R3: User access mechanisms for HPC resources

R4: Training facilities, documentation, and other information resources

R5: Coordinated tri-lab user support functions

Gaps

G1: Resources enabling equal partnership with LANL for ACES user support

G2: Resources to fulfill expanding user support needs associated with the NSCC

G3: Limited self-help resources (for end-users and user support personnel)

G4: Lack of familiarity with Sequoia and future ASC systems, including details related to evolving architectures and associated software environments

G5: Collaborative tools for enhancing user support productivity and effectiveness, within local environments and across the tri-lab (and broader) HPC community

Five-Year Plan

FY14

- Provide user support for SNL and tri-lab ASC computing ~~systems~~
- Deliver user support for Sequoia (G4)
- Continue ~~efforts~~ to improve collaborative tools and self-help resources, particularly in support of ACES (G5)
- Continue to partner with LANL to strengthen joint user support for ACES platforms; prepare to provide Trinity user support (G1, G4)
- Begin addressing user support needs for the National Security Computing Center (NSCC) (G2)
- Continue to leverage Information Technology Infrastructure Library (ITIL) as a framework for improving the HPC OneStop Service Desk processes and practices
- ~~Explore expanding and/or migrating services into the iHPC to support tri-lab collaborative user support activities~~

FY15

- Provide user support for Sandia and tri-lab ASC computing systems
- Develop expertise in support of next-generation architectures and software environments (G4)
- Ramp up to deliver user support for Trinity (G4)
- Fully deploy user support for NSCC (G2)

FY16

- Provide user support for Sandia and tri-lab ASC computing systems
- Deliver user support for next-generation computing environments (G4)

FY17

- Provide user support for Sandia and tri-lab ASC computing systems
- Continue enhancement of user support resources and capabilities, leveraging trends in new technologies (G3, G5)

FY18

- Provide user support for Sandia and tri-lab ASC computing systems
- Continue enhancement of user support resources and capabilities, leveraging trends in new technologies (G3, G5)

Collaborations (WBS 1.5.5.3)

This level 4 product provides programmatic support for collaboration with external agencies on specific HPC projects. This product also includes collaborations with internal or external groups that enable the program to improve its planning and execution of its mission.

Program Support (LLNL)

The Program Support project provides service to the ASC program. Program Support services include procurement and contracting, project management, and meeting support. These services are in support of both tri-lab and LLNL-only activities, including collaborations with academic, industrial, and other government agencies.

Required Capabilities

R1: Support procurement, planning, collaboration, meetings, and contracting needs of NNSA tri-lab ASC program

Gaps

None

Five-Year Plan

FY14

- Continue FY13 procurement, contract management, and extreme-scale computing ~~program~~ planning ~~needs~~
- Support annual HPC Best Practices workshop with Office of Science
- Support bi-annual Predictive Science Panel (PSP) meetings
- Support PCASE awardee
- Manage new ~~Support start of~~ PSAAP2 program
- ~~Plan TLCC3 tri-lab procurement and vendor selection process~~
- ~~Continue CD process for 2017 Advanced Technology System~~

FY15

- Continue FY14 procurement, contract management, and program planning needs
- Support annual HPC Operations Review meeting with Office of Science
- Support bi-annual Predictive Science Panel (PSP) meetings
- Manage CTS tri-lab procurement planning and vendor selection process, and initial system deliveries
- ~~Continue contract management for delivery of TLCC3 systems to tri-labs~~
- Manage ~~Continue~~ CD process and HQ reporting for the Sierra ATS ~~for 2017 Advanced Technology System~~

FY16

- Continue FY15 procurement, contract management, and program planning needs
- Support annual HPC Operations Review meeting with Office of Science

- Support bi-annual PSP meetings
- Manage ~~Continue contract management for~~ delivery of CTS systems to tri-labs
- Manage ~~Continue~~ CD process for 2017 Advanced Technology System

FY17

- Continue FY16 procurement, contract management, and program planning needs
- Continue contract management for delivery of CTSto tri-labs
- Manage ~~Continue~~ CD process and HQ reporting for the Sierra ATS, ~~for 2017 Advanced Technology System,~~ including selection and system delivery

FY18

- Continue FY17 procurement, contract management, and program planning needs
- Support annual HPC Operations Review meeting with Office of Science
- Support bi-annual PSP meetings
- Write the Sierra ATS CD4 and final Sierra ATS HQ reports

Program Support (LANL)

Through the Program Support project, LANL provides support to the national program, both by providing resources and expertise to the Federal program office and by participating in coordination and integration activities for the tri-lab program.

Required Capabilities

R1: Support ASC Principal Investigator (PI) meeting

R2: Host Predictive Science Panel (PSP) meeting

R3: Publish ASC eNews online newsletter

R4: Support PSAPP

R5: Tri-lab publications

Gaps

None

Five-Year Plan

FY14

- Participate in Principle Investigator and PSP Meetings
- Publish ASC eNews online newsletter

FY15

- Host PSP Meeting
- Publish ASC eNews online newsletter

FY16

- Participate in PI and PSP meetings
- Publish ASC eNews online newsletter

FY17

- Host PSP Meeting
- Publish ASC eNews online newsletter

FY18

- Participate in PI and PSP meetings
- Publish ASC eNews online newsletter

Program Support (SNL)

The Program Support project provides critical coordination and integration activities essential to the success of ASC. It is divided into two distinct parts: 1) provide ASC programmatic planning, reviews, and communications; and 2) provide SNL outreach to the other institutions and programs.

This capability is critical to the ASC SNL program integration, communication, and management within the laboratory and with the external community. A significant management and integration function in this project is captured in the SAIC contract that provides support for NNSA Headquarters and SNL in communications and logistics. External advisory boards supported through this project also provide feedback to the ASC leadership team regarding the maturation of the predictive engineering sciences capability and the quality of SNL's computational science R&D. Support of external collaborations, including the University Alliance program and the exascale initiative (with DOE/SC), is also included in this project.

Required Capabilities

R1: Management of ASC program cost, schedule, and performance

R2: Management of SAIC to provide administration support for ASC HQ

R3: External reviews of program-supported predictive capability development

R4: Engagement with ASC federal managers and tri-lab ASC executives

R5: Technical engagement and program planning for the Predictive Science Academic Alliances Program (PSAAP)

R6: Technical engagement and program planning for the proposed DOE Exascale Initiative

Gaps

G1: Yearly Predictive Engineering Sciences Panel Reviews

G2: Yearly support for external reviews hosted by SNL organizations implementing predictive simulation work

G3: Yearly support for ASC program management, including interaction with federal managers and tri-lab executives, and the proposed DOE Exascale Initiative

G4: Yearly management support of the SAIC contract

G5: Yearly support of PSAAP activities

Five-Year Plan

FY14

- Organize and host fifth Predictive Engineering Science Panel (PSEP) meeting and side meetings of PESP sub-panels (G1)

- Support external review panel meetings for QASPR, the Engineering Sciences External Advisory Board, and the Computational Sciences External Advisory Board (G2)
- Support programmatic needs of the PSAAP-2 program and the proposed DOE Exascale Initiative (G3, G5)
- Manage the SAIC contract to provide various administration support for HQ (G4)
- Support programmatic needs of NNSA tri-lab ASC program and ASC executive committee (G3)

FY15

- Organize and host sixth Predictive Engineering Science Panel meeting (G1)
- Support external review panel meetings for QASPR, the Engineering Sciences External Advisory Board, and the Computational Sciences External Advisory Board (G2)
- Support programmatic needs of the PSAAP II program ~~and the proposed DOE Exascale Initiative~~ (G3, G5)
- Manage the SAIC contract to provide various administration support for HQ (G4)
- Support programmatic needs of NNSA tri-lab ASC program and ASC executive committee (G3)

FY16

- Organize and host seventh Predictive Engineering Science Panel meeting (G1)
- Support external review panel meetings for QASPR, the Engineering Sciences External Advisory Board, and the Computational Sciences External Advisory Board (G2)
- Support programmatic needs of the PSAAP II program ~~and the proposed DOE Exascale Initiative~~ (G3, G5)
- Manage the SAIC contract to provide various administration support for HQ (G4)
- Support programmatic needs of NNSA tri-lab ASC program and ASC executive committee (G3)

FY17

- Organize and host eighth Predictive Engineering Science Panel meeting (G1)
- Support external review panel meetings for QASPR, the Engineering Sciences External Advisory Board, and the Computational Sciences External Advisory Board (G2)
- Support programmatic needs of the PSAAP II program ~~and the proposed DOE Exascale Initiative~~ (G3, G5)

- Manage the SAIC contract to provide various administration support for HQ (G4)
- Support programmatic needs of NNSA tri-lab ASC program and ASC executive committee (G3)

FY18

- Organize and host ninth Predictive Engineering Science Panel meeting (G1)
- Support external review panel meetings for QASPR, the Engineering Sciences External Advisory Board, and the Computational Sciences External Advisory Board (G2)
- Support programmatic needs of the PSAAP II program
- Manage the SAIC contract to provide various administration support for HQ (G4)
- Support programmatic needs of NNSA tri-lab ASC program and ASC executive committee (G3)

Applications in Support of Manufacturing Production and Connectivity (Y-12)

The Applications in Support of Manufacturing Production and Connectivity project supports the utilization of ASC codes and computing resources to solve production manufacturing problems through modeling and simulation. The project includes support for connecting to ASC computing resources and job submission, execution, and visualization. The project provides the infrastructure necessary to test applications and scenarios before deployment on larger ASC resources. Development and deployment of software to support the solution of manufacturing problems is also supported by the project. Visualization techniques that can be utilized in the Y-12 network and computing infrastructure will be evaluated and implemented. Finally, participation in Nuclear Weapons Complex ASC-related activities is covered.

Required Capabilities

R1: Local cluster computer capable of code development and performing small to mid-scale analysis of radiation transport, heat transfer, and image, and signal analysis problems

R2: Access to commercial and ASC-developed software for heat transfer, fluid flow, and solid mechanic (finite-element) analysis

R3: Support for development of models to solve production manufacturing problems

Gaps

G1: Aging cluster computing capability with operating system that limits software codes that can be installed in support of manufacturing analysis

G2: Limited access to GPU-based architectures

G3: Enhanced training environment does not exist for hazardous manufacturing operations

Five-Year Plan

FY14

- Configure and demonstrate an initial virtual-reality training environment for hazardous manufacturing operations; integrate a physical glove-box interface with an immersive stereoscopic computer simulation and demonstrate the applicability for hazardous operations training
- Upgrade immersive environment software and hardware; evaluate advanced optical tracking systems and haptic feedback solutions
- Work with the Uranium Processing Facility Process Engineering staff to develop an ergonomic evaluation capability to certify process equipment designs based on real-time motion tracking to control a computer-generated ergonomic manikin simulation

- Enhance parallel cluster environment to include COMSOL, QMU for criticality excursions, adjoint-forward ATTILA for multiple particle nuclear radiation source term forensics, and MCNP6.1
- Continue to explore the use of GPU architectures to accelerate image and signal processing tasks (this will require use of JICS and/or SNL resources, with the local Linux cluster used as a test bed)
- Continue to evaluate new codes on the Y-12 cluster and utilize Y-12 and remote ASC cluster resources to solve production manufacturing problems
- Participate in National Security Enterprise ASC activities
- ~~Continue to evaluate new versions of SIERRA codes when available with emphasis to improve manufacturing and analysis capabilities at Y-12~~
- ~~Continue to collaborate with other sites (NWC colleagues) to best utilize codes and deploy capabilities as necessary to enable collaboration~~
- ~~Analyze modal data using SALINAS and high-fidelity component models~~
- ~~Continue to evaluate new codes on the Y-12 cluster and utilize Y-12 and remote ASC cluster resources to solve production manufacturing problems (G1, G2)~~

FY15

- Upgrade the stereoscopic immersive environment hardware and software
- Implement a gesture-based capability to view real-time procedures during an immersive training exercise
- ~~Investigate use of UQ for large-scale inverse problems of interest to Y-12 (for example, image restoration and reconstruction of tomographic data on selected stockpile components)~~
- Continue to evaluate new codes on the Y-12 cluster and utilize Y-12 and remote ASC cluster resources to solve production manufacturing problems
- Participate in National Security Enterprise ASC activities

FY16

- Implement the capability for team participation in virtual reality training scenarios
- Upgrade simulation modeling computing infrastructure to support manufacturing simulation and production modeling
- Continue to collaborate with other sites (NWC colleagues) to best utilize codes and deploy capabilities as necessary to enable collaboration
- Continue to evaluate new codes on the Y-12 cluster and utilize Y-12 and remote ASC cluster resources to solve production manufacturing problems
- Participate in National Security Enterprise ASC activities

FY17

- Continue to collaborate with other sites (NWC colleagues) to best utilize codes and deploy capabilities as necessary to enable collaboration
- Continue to evaluate new codes on the Y-12 cluster and utilize Y-12 and remote ASC cluster resources to solve production manufacturing problems
- Participate in National Security Enterprise ASC activities

FY18

- Continue to collaborate with other sites (NWC colleagues) to best utilize codes and deploy capabilities as necessary to enable collaboration
- Continue to evaluate new codes on the Y-12 cluster and utilize Y-12 and remote ASC cluster resources to solve production manufacturing problems
- Participate in National Security Enterprise ASC activities

System and Environment Administration and Operations (WBS 1.5.5.4)

This level 4 product provides necessary operational support for reliable production computing and storage environments. The following activities are included: system administration and operations, software and hardware maintenance, licenses and contracts, computing environment security and infrastructure, requirements planning, initial deployment, production computing services, and tri-lab system integration and support.

System and Environment Administration and Operations (LLNL)

This project provides necessary operational support for reliable production computing environments. The following activities are included: system administration and operations, software and hardware maintenance, licenses and contracts, computing environment security and infrastructure, requirements planning, initial deployment, production computing services, and tri-lab system integration and support. Included within the scope of this product is the operational support for systems used as part of partnerships with academic, industrial, and other governmental agencies.

Required Capabilities

R1: System configuration

R2: System and user security development and support

R3: System administration and monitoring

R4: File system support

R5: 24x7 system monitoring and user support

R6: Self hardware maintenance of all platforms

R7: Logistics coordination and spares management

Gaps

G1: Reduced staffing requires continued focus on efficient support, automation, and process improvement

~~None~~

Five-Year Plan

FY14

System Administration and Operations:

- Dismantle and dispose of udawn, rzdawndev, and dawn
- Deploy Sequoia visualization cluster Max
- Deploy expanded lustre file system hardware (marzen, porter, stout) in advance of ZFS-based software transition
- Retire Coastal and Juno
- Continue 24 x 7 x 365 monitoring and diagnostics of facility and systems
- Upgrade Weapons & Complex Integration's test and development cluster on the open computing facility (replace rzalastor)
- Integrate Splunk into daily monitoring
- Replace NFS home directory servers

- Replace /nfs/tmp servers

Security Technology:

- Deploy full Identity Management capability (classified accounts) on SCF by replacing IAM
- Upgrade to RSA 8 in the secure computing facility and SNSI environments
- Consolidate infrastructure systems to a virtual machine environment
- Implement log-based security event analysis and detection
- ~~Retire Dawn~~
- ~~Perform ongoing system administration and operation of production platforms and file systems~~
- ~~Perform ongoing system and user security development and support~~

FY15

- Deploy CTS-1 ~~TLCC3~~
- Retire TLCC1 systems
- Perform ongoing system administration and operation of production platforms and file systems
- Perform ongoing system and user security development and support

FY16

- Perform ongoing system administration and operation of production platforms and file systems
- Perform ongoing system and user security development and support

FY17

- Deploy 2017 Advanced Technology System
- Perform ongoing system administration and operation of production platforms and file systems
- Perform ongoing system and user security development and support

FY18

- Perform ongoing system administration and operation of production platforms and file systems
- Perform ongoing system and user security development and support

System Administration and Storage (LANL)

The System Administration and Storage project covers all services for computational systems operated by LANL for the purpose of providing an HPC production computing environment for weapons designers, developers, and engineers. The project works with users to troubleshoot problems experienced while running their applications, and helps users transition from old to new computing platforms. The capabilities include system configuration, system and user security, resource management, system administration and monitoring, archival storage, Panasas, and NFS file systems.

Required Capabilities

R1: System configuration

R2: System and user security

R3: Resource management

R4: System administration and monitoring

R5: Archival storage

R6: File system support

Gaps

G1: Adequate expertise shortage due to attrition and early retirements; expertise needs to be hired or cultivated

Five-Year Plan

FY14

- ~~Retire Roadrunner~~
- Support HPC systems by conducting ongoing and daily system and storage administration with continuous monitoring of production systems and infrastructure servers
- Ensure workload is carried out by proper configuration of queues and scheduling policies plus daily monitoring and problem resolution relating to workloads running on HPC computing resources
- Provide direct user support for ASC/HPC systems/architectures through in-person consulting and online documentation (G1)
- Expand ~~production~~ Lustre file system (G1)
- Integrate Zenoss-based HPC monitoring into HPC facility monitoring infrastructure
- Implement software for upgrade of current monitoring system to latest open-source Zenoss version

FY15

- Provide support for Trinity site preparation project
- Retire Typhoon and Cielo
- Provide support for ASC/HPC systems/architectures
- Provide support for CTS-1 integration and production readiness

FY16

- Retire Luna, Moonlight, Typhoon, Cielo
- Provide support for ASC/HPC systems/architectures
- Provide support for Trinity integration readiness milestone
- Provide support for CTS-1 integration and production readiness

FY17

- Provide support for ASC/HPC systems/architectures
- Provide support for Trinity production readiness milestone

FY18

- Provide support for ASC/HPC systems/architectures

Operations and Procurement Support (LANL)

The Operations and Procurement Support project provides around-the-clock operations and monitoring of the scientific computing resources, including performance computers such as ~~Roadrunner, Hurricane,~~ Luna, ~~Lobo,~~ Typhoon, ViewMaster, Cielo, ~~Mapache,~~ Moonlight, and data storage and retrieval systems such as the HPSS. In addition to monitoring all components 24x7, the computer operators provide systems hardware maintenance for all ASC platforms. Working with the vendor system engineers, the operators also provide backup hardware support for the Cielo capability system. This includes all components of the production computing environment, from compute engines, hardware, fileservers, archival storage systems, the facilities they reside in and utilities they are dependent upon, to all required software on these systems.

The procurement support aspect of this project assists customers with the technical and administrative aspects of planning, procurement, and contractual agreements for computer hardware and software products and services.

Required Capabilities

R1: Provide 24x7 operations and monitoring of ASC/HPC resources

R2: Provide hardware maintenance for ASC/HPC resources

R3: Provide procurement support for ASC/HPC hardware, software, services, and contracts

Gaps

G1: Adequate expertise shortage due to attrition and early retirements; expertise needs to be hired or cultivated

G2: Partner with vendors to enable LANL to perform self-maintenance on ASC/ACES platforms

Five-Year Plan

FY14

- Provide 24x7 operations and monitoring of HPC computing resources
- Provide hardware self-maintenance for current and future ASC platforms (G2)
- Develop requirements for upgraded tools for monitoring HPC platforms and file systems
- Provide support for SCC upgrades for Trinity system.
- ~~Provide system administration, network administration and filter development on monitoring and hardware testing infrastructure (G1)~~
- Provide technical and administrative support for procurement of HPC platforms, supporting hardware and software, and other products and services required by HPC

FY15

- Provide 24x7 operations and monitoring of HPC computing resources
- Provide hardware self-maintenance for current and future ASC platforms (G2)
- Decommission ~~Hurricane,~~ Cielo, and Cielito
- Provide system administration, network administration, and filter development on monitoring and hardware testing infrastructure (G1)
- Provide technical and administrative support for procurement of HPC platforms

FY16

- Provide 24x7 operations and monitoring of HPC computing resources
- Provide hardware self-maintenance for current and future ASC platforms (G2)
- Decommission Luna, Moonlight and Typhoon ASC Platforms
- Provide system administration, network administration, and Zenoss interface development on monitoring and hardware testing infrastructure
- Provide technical and administrative support for procurement of HPC platforms

FY17

- Provide 24x7 operations and monitoring of HPC computing resources
- Provide hardware self-maintenance for current and future ASC platforms
- Provide system administration, network administration, and Zenoss interface

FY18

- Provide 24x7 operations and monitoring of HPC computing resources
- Provide hardware self-maintenance for current and future ASC platforms
- Provide system administration, network administration, and Zenoss interface
- Provide technical and administrative support for procurement of HPC platforms

~~*Requirements Planning (LANL)*~~

~~The Requirements Planning project collects and understands user requirements for production computing resources and quality of service, and to develop new metrics, data collection, and analysis techniques. The project is currently focused on establishing dependencies between cluster availability and cluster utilization and on implementing new technologies to accommodate rapidly growing information sets.~~

~~*Required Capabilities*~~

~~R1: Data collection from HPC computing resources~~

~~R2: Analysis and reporting on resource availability and utilization~~

~~*Gaps*~~

~~G1: Adequate expertise shortage due to attrition and early retirements; expertise needs to be hired or cultivated~~

~~*Five-Year Plan*~~

~~FY14~~

- ~~Integrate Zenoss-based HPC monitoring into HPC facility monitoring infrastructure~~
- ~~Implement software for upgrade of current monitoring system to latest open-source Zenoss version~~

~~FY15~~

- ~~Improve reporting capabilities on monitoring and data collection suite of tools~~

~~FY16~~

- ~~Enhance data collection, reporting, and troubleshooting suite of tools~~

~~FY17~~

- ~~Enhance data collection, reporting, and troubleshooting suite of tools~~

Computing Platform Integrations and Deployment (LANL)

The scope of the Computing Platform Integration and Deployment project is to accept delivery and begin deployment of production TLCC and CTS. This includes participating in developing the design requirements as part of a tri-lab requirements planning team. Primary capabilities include completing the acceptance tests, diagnostics test, integrating the systems into the LANL unclassified network, system stabilization, and transition into the classified network. Included in this project is support for the ASC Commodity Technology System acquisition strategy and provision for requirements that help to achieve the strategy.

The objective of the project is the integration of all hardware and software components to deliver a system environment to application users for programmatic work. This includes site preparation to prepare the Simulation and Computing Complex (SCC) facility for deploying these production capacity systems. The integration and deployment activities will focus on the following areas: System/Operating System, File Systems, Interconnect, External Network including PaScalBB, Regression Testing, Monitoring, and Application Readiness.

Required Capabilities

R1: Platform acceptance and diagnostic testing

R2: Platform integration into LANL unclassified and classified networks

R3: System stabilization

Gaps

G1: Continued definition of requirements and workload justifications for Commodity Technology Systems

G2: Ensuring that the balance of the Commodity Technology Systems include file systems, networking, and archival storage for the continued production operation of these capacity production systems for the LANL applications community

Five-Year Plan

FY14

- Continue to update the system environment for capacity systems (G2)
- Develop and coordinate the site preparation requirements for CTS-1 (G2)

FY15

- Deploy the CTS-1 at Los Alamos
- Develop plan and schedule for system accreditation

FY16

- Provide production support for CTS-1 system

- ~~Develop requirements, with tri-lab design team, for Commodity Technology System FY17 (G1, G2)~~
- ~~Deploy and integrate Commodity Technology System 2~~

FY17

- Provide production support for CTS-1 ~~TLCC3~~ systems

FY18

- Develop requirements, with tri-lab design team, for CTS-2 (G1, G2)
- Provide production support for CTS-1 systems

Production Computing Services (SNL)

The Production Computing Services project's goals are to operate and maintain all ASC production platforms and associated support systems, and operate data services and visualization systems, long-term hierarchical storage services, high-performance network systems, tri-lab compatible cyber authentication and authorization systems, and monitoring and reporting services. This project supports tri-lab capability platform resource allocations and coordinates with tri-lab peers in establishing priority scheduling, if required. This project coordinates the integration and deployment of TLCC capacity systems into SNL's production computing environment, in collaboration with WBS 1.5.5.6 Common Computing Environment. Support of CCE common service and environment decisions and configuration management activities are also provided.

This project has expertise in operating capacity computing clusters; integrating file servers at the system or facility-wide level; deploying new computing, storage, and data management platforms; and in retiring end-of-life platforms. System administration for complex HPC environments is provided, as are design and development activities for new innovative computing platforms.

Required Capabilities

R1: Computing facility space, power, and cooling sufficient for current and future platforms

R2: Experienced technical staff and support personnel versed in computing systems (both hardware and software), networks and protocols (interconnects and local networks), file systems and batch management

R3: Wide Area High Speed network design and operations support

R4: Data transfer and storage systems and support personnel

R5: Tier 2 and Tier 3 support for trouble resolution

Gaps

G1: Aging facility limits agility in accepting new high-density computer platforms

G2: Staffing levels are insufficient due to recent losses and the aging demographic of existing staff; few entry-level personnel in HPC are on role to begin training as replacements

G3: Limited workspace for analysts and system administrators creates inefficiencies in utilizing the computing resources; larger collaborative spaces that support a larger team in the SCI environment would improve productivity

G4. Insufficient compute capacity to support full-system and full-resolution analysis

G5. Data separation and integrity technologies are inadequate for some potential customers of NSCC resources

Five-Year Plan

FY14

- Continue operations support for Cielo, RFP activity on Trinity, and coordination with LLNL on Sequoia software stack needs for SNL mini-Sequoia system
- Continue operations of storage systems, archive systems, and production systems supporting ASC and nuclear weapons programs
- Deploy production version of LDMS on SNL internal and collaborative partners large-scale production HPC platforms
- Develop Red Hat package manager (RPM) for TOSS to include LDMS
- Continue operations of storage systems, archive systems, and production systems supporting the NSCC programs
- Provide additional resources for PSAAP program use in the Open HPC network
- Acquire a flexible computing platform for the NSCC to provide capacity computing and, when needed, expanded capability computing (G4)
- ~~Design new computing facility (G3)~~
- ~~Expand NSCC capacity resources (G4)~~
- ~~Prepare acquisition plan for large NSCC resource to be obtained in FY15–FY16 (G4)~~
- ~~Retire first generation capacity compute clusters TLCC (G1)~~
- ~~Continue ACES engagement and support of Trinity system procurement (G4)~~

FY15

- Construct plans for new computing facility (G3)
- Begin procurement activity for next-generation capacity systems (G4)
- Initiate procurement activity for NSCC computing resource (G4)
- Perform ACES integration activities for Trinity; engage with NERSC
- Continue operations of all production systems

FY16

- Complete installation and acceptance of Capability Class NSCC resource (G4)
- Integrate expansion of file systems and data archive platforms sized to service the Capability Class system (G5)
- Continue operations of all production systems

FY17

- Integrate next generation of capacity compute platforms into ASC and NSCC environments CTS-1 ~~TLCC3~~ (G4)

- Continue operations of all production systems

FY18

- Continue operations of all production systems

Facilities, Network, and Power (WBS 1.5.5.5)

This level 4 product provides necessary physical facility and other utility infrastructure. The following activities are included: facilities infrastructure, classified and unclassified facility networks, wide-area classified networks, ongoing network operations, infrastructure integration, and power.

Facilities, Network, and Power (LLNL)

The Facilities, Network, and Power project provides for the necessary physical facilities, utilities, and power capabilities to ASC systems. Work in this area includes adequate raised floor space, flexible cooling solutions, and power to site large-scale ASC platforms. In addition, this project funds needed office, meeting room, and auxiliary space to enable a highly motivated and effective staff. Also included are classified and unclassified facility networks, wide-area classified networks, and ongoing network operations. This project also enables enhanced collaborations with academic and industrial partners.

Required Capabilities

R1: Provide support for electrical, mechanical, cooling, and network services infrastructures

R2: Manage ASC computing facility upgrades

R3: Support network backbones, hardware and software, in LLNL CZ, RZ, and SCF networks

R4: Support DisCom WAN, iHPC configurations and ESNet

Gaps

G1: IB network distance limitation as compared to Ethernet; current strategy is to move from 40 G to 100 G Ethernet between buildings

G2: Limited tools for IB network monitoring/management and configuration

G3: Need ability to monitor and synthesize all facilities/networks/systems data

G4: Aging building 451 is increasing facilities maintenance costs and limits unclassified expansion capabilities at reasonable cost because of power limitations

Five-Year Plan

FY14

- Begin construction of the new unclassified HPC facility to house unclassified systems, including the next-generation CTS-1 clusters
- Begin a project to bring in electrical equipment that allows for more widely varying voltages in preparation for the 2017 system
- Complete the cutover of the B453 chilled-water system small staging chiller with a larger in-line chiller to increase reliability and reduce nuisance staging of the chilled water system
- Continue to enhance diagnostic and monitoring of Infiniband (IB) fabrics on Sequoia and other IB-attached Lustre file systems
- Continue to analyze and evaluate emerging network technologies

- ~~Provide ongoing operations and maintenance of electrical and mechanical systems for ASC computing facilities~~
- ~~Complete construction of 6000-sq.ft., unclassified modular building (G4)~~
- ~~Evaluate viability of using 40/100 G Ethernet for core connectivity~~
- ~~Evaluate use of EDR for file system networks~~
- ~~Provide design, technical support, configuration management, technical deployment expertise for all system and components of the LC's networking infrastructure~~

FY15

- Provide ongoing operations and maintenance of electrical and mechanical systems for ASC computing facilities
- Complete construction of ~~Commission~~ new unclassified modular building and extend electrical and mechanical infrastructure for new systems
- Provide design, technical support, configuration management, technical deployment expertise for all system and components of the LC's networking infrastructure

FY16

- Provide ongoing operations and maintenance of electrical and mechanical systems for ASC computing facilities
- Begin Advanced Technology System facilities preparation project
- Provide design, technical support, configuration management, technical deployment expertise for all system and components of the LC's networking infrastructure

FY17

- Commission facilities for 2017 Advanced Technology System
- Provide ongoing operations and maintenance of electrical and mechanical systems for ASC computing facilities
- Provide design, technical support, configuration management, technical deployment expertise for all system and components of the LC's networking infrastructure

FY18

- Provide ongoing operations and maintenance of electrical and mechanical systems for ASC computing facilities
- Provide design, technical support, configuration management, and technical deployment expertise for all system and components of the LC's networking infrastructure

Facilities, Networking, and Power (LANL)

The Facilities, Networking, and Power project is responsible for the engineering, design, operation, and maintenance of the mission-important electrical, mechanical, cooling, network services, and other computing infrastructure in support of the ASC program. The project provides support for infrastructure design upgrades, project and space management, user interface and oversight, demolition and decommissioning of older systems, network backbones, user LANs, classified/unclassified network hardware and services, DisCom WAN, and computer site preparation for new platforms. Because the tri-lab community requires the systems to be operational at all times, the project provides on-call support after hours and on weekends for facility related issues.

Required Capabilities

R1: Provide support for electrical, mechanical, cooling, and network services infrastructures

R2: Manage ASC computing facility upgrades

R3: Provide decommissioning services for resources reaching end-of-life

R4: Support network backbones, hardware and software, in LANL yellow and red networks

R5: Support DisCom WAN

Gaps

G1: The SCC power upgrade is dependent on the completion of the 115-KV line project that is described in the LANL Master Power Plan

G2: Adequate resources to support the increased demand for data movement on the inter- and-intra-laboratory network and archive infrastructure

Five-Year Plan

FY14

- Complete construction of SCC Trinity Infrastructure Project
- ~~Complete mechanical and electrical site preparation for Trinity Platform~~
- ~~Complete mechanical and electrical site preparation for CTS-1 Capacity Computing~~
- Provide ongoing operations and maintenance of electrical and mechanical systems for ASC computing facilities
- Publish design document for Next-Generation BackBone
- ~~Final design for medium voltage power project for ATS-3 Platform~~
- ~~Resources identified in support of data movement (G2)~~

FY15

- Provide ongoing operations and maintenance of electrical and mechanical systems for ASC computing facilities
- Conceptual design for ATS-3 Platform
- Data movement resources integrated into LANL secure network (G2)
- Start NGBB build and integration

FY16

- Provide ongoing operations and maintenance of electrical and mechanical systems for ASC computing facilities
- Design SCC upgrade for mechanical and electrical infrastructure project for ATS-3 (G1)
- ~~Perform design for mechanical and electrical infrastructure project for ATS-3~~ ~~(G1)~~
- ~~Complete construction for medium voltage power project for ATS-3 platform~~ ~~(G1)~~
- Complete NGBB build and integration

FY17

- Provide ongoing operations and maintenance of electrical and mechanical systems for ASC computing facilities
- Perform construction of SCC ATS-3 infrastructure project (G1)
- ~~Complete mechanical and electrical site preparation for ATS-3~~
- ~~Completion of 115-KV power line (G1)~~

FY18

- Provide ongoing operations and maintenance of electrical and mechanical systems for ASC computing facilities
- ~~Complete mechanical and electrical site preparation for ATS-3 (G1)~~
- Perform construction of SCC ATS-3 infrastructure project (G1)

Facilities, Networking, and Power (SNL)

The Facilities, Networking, and Power project funds the power and space charges assigned to HPC systems (capacity and file system servers) and long-term hierarchical storage servers (running the HPSS software product). It provides for facilities and personnel to manage installation and removal of computing platforms, file systems, visualization systems, networking equipment, power distribution systems, and cooling systems in support of all computing resources. It also funds major operations contracts such as the ASC Distance Computing (DISCOM) WAN.

Facilities professionals have reduced overall operating expenses by minimizing cooling and electrical distribution expenses over the last several years through a comprehensive program of introducing more efficient computer room air conditioning units, using higher voltage electrical source power distribution units, exploring alternative energy sources and conservation mechanisms, which include reducing the volume of chilled water required for cooling and improving air flow in the facility by minimizing obstructions underneath the computer floor. These efforts have been recognized with several SNL-specific and national awards, including three 2011 EStar Awards from the DOE Office of Sustainability Support.

Required Capabilities

R1: Facilities, power, and cooling expertise needed to continue energy conservation practices and inform new facility designs for high-efficiency energy use

R2: Network expertise and continued operational funding for DisCom WAN linking LANL, LLNL, and SNL campuses

Gaps

G1: Aging facility limits agility in accepting new high-density computer platforms

G2: Limited workspace for analysts and system administrators creates inefficiencies in utilizing the computing resources; larger collaborative spaces that support a larger team in the SCI environment would improve productivity

G3: Cost of electricity and cooling continue to be a significant portion of the operating expenses for systems

Five-Year Plan

FY14

- Upgrade SNL's two 10-GE links to LANL to a single 100-G link
- Deploy two 10-GE encryptors at each site to double the aggregate available bandwidth
- Manage operation of the DisCom WAN
- Renew planning activities for Building 725 expansion options
- ~~Begin expansion of NSCC computer facility and analyst workspace (G2)~~

FY15

- Prepare facility for integration of capability-class computer system for NSCC (G1)
- Analyze networking technologies for next-generation upgrade of DisCom WAN
- Expand Building 725 computer floor space (G1)

FY16

- Configure power and cooling, and manage installation efforts for capability computing platform in expanded NSCC facility (G1)

FY17

- Introduce new technologies to DisCom WAN
- Prepare facility for next-generation capacity systems

FY18

- Continue energy saving practices integrating new cooling technologies throughout all facilities

Common Computing Environment (WBS 1.5.5.6)

The goal of the CCE product is to enable a common environment across the tri-labs that will initially be deployed on the TLCC systems. The scope of this product includes funded R&D projects to address gap areas identified by the tri-lab technical working groups.

The CCE working groups and projects focus on a common software stack, including but not be limited to, OS software; application development tools; resource management; HPC monitoring and metrics; and common tri-lab environment issues such as configuration management, licenses, WAN access, and multi-realm security.

System Software Deployment for Commodity Technology Systems

The Tri-Lab Operating System (TOSS) is the software stack that runs on Linux capacity clusters, initiating with TLCC platforms delivered in FY08. The goal of the TOSS project is to increase efficiencies in the ASC tri-lab community with respect to both the utility and the cost of the CCE. A notable achievement of this effort has been the porting of TOSS to SNL's RedSky system, given that its architecture more closely resembles many current capability systems rather than the typical commodity-based Linux capacity cluster.

This project delivers a fully functional cluster OS capable of running MPI jobs at scale on TLCC hardware. The system must meet CCE requirements for providing a common software environment on TLCC hardware across the tri-lab complex, now and into the future.

TOSS provides a complete product with full lifecycle support. Well-defined processes for release management, packaging, quality assurance testing, configuration management, and bug tracking are used to ensure a production-quality software environment can be deployed across the tri-lab in a consistent and manageable fashion.

Required Capabilities

R1: Fully functional cluster operating system (kernel, Linux distribution, IB stack and related libraries, and resource manager)

R2: Capable of running MPI jobs at scale on Linux capacity clusters

R3: Full lifecycle support including release management, packaging, QA testing, configuration management, and bug tracking

Gaps

G1: CTS-1 ~~TLCC3~~ hardware support/integration unknown

G2: Integration with Red Hat release schedule and evolutionary changes

G3: Pace of development of GPU-like processors requires increased investigation of these processors and development of solutions to integrate these into advanced systems

G4: SLURM is open source but no longer a solely DOE-funded project; tri-lab-specific requirements may develop that are not relevant or important to the worldwide SLURM community

Five-Year Plan

FY14

- Provide ongoing TOSS software development and support (G2)
- Develop/deploy TOSS 2.X (based on RHEL 6.X) (G2)
- ~~Investigate new generation of GPU-like processors, including MICs and Fusion (G3)~~

- ~~Deploy the next generation of the ASC TLCC systems (TLCC3), which includes software integration and testing for the tri-lab environment (G1, G2, G3)~~
- Initiate development of TOSS 3.X (based on RHEL 7.X) (G2)
- Continue SLURM support efforts through tri-lab collaboration
- Develop identified collaborative system software tasks, including investigation of accelerator architectures; GPGPUs, MICs and Fusion; integration of virtualization; logging/monitoring improvements; and testing infrastructure improvements

FY15

- Provide ongoing TOSS software development and support (G2)
- Develop/deploy TOSS 3.X (based on RHEL 7.X) (G2)
- Continued assessment and integration of new architecture into TOSS capability (G3)

FY16

- Provide ongoing TOSS software development and support (G2)
- Continued assessment and integration of new architecture into TOSS capability (G3)

FY17

- Provide ongoing TOSS software development and support (G2)
- Continued assessment and integration of new architecture into TOSS capability (G3)

FY18

- Provide ongoing TOSS software development and support (G2)
- Continued assessment and integration of new architecture into TOSS capability (G3)

Programming Environment Development/Support for Tri-Lab Systems

The goals of the Programming Environment Project are to enhance productivity of the tri-lab application development teams, operation teams, and analysts. This project achieves these goals by developing and deploying user tools and programming environments to support a variety of applications running on tri-lab HPC resources. It leverages skills across the tri-labs to meet the challenges posed by rapid changes in processor and systems technology and evolving programming models. This project entails software integration, feature enhancements, installation, training, support for vendor provided tools, open source software tools, and lab-developed tools.

Open|SpeedShop (O|SS) is a tri-lab supported open source project that provides a wide range of performance experiments within a single environment. This includes support for PC sampling, inclusive and exclusive execution times for routines, hardware counters, as well as I/O and MPI tracing. O|SS is designed to work on binaries of application without the need to recompile, enabling a clean and easy integration into the development workflow. Transition to the Component Based Tool Framework (CBTF), which is designed to facilitate easy addition of new capabilities, is currently underway.

Debuggers are another key effort within this project. All three labs are utilizing TotalView as the core debugger. LLNL has the more strategic relationship with Rogue Wave Software and is also doing additional research into other debugging tools (for example, STAT). The approach to subset debugging is key as ASC moves toward larger scales and collects information that helps focus in on the trouble area. Input from all three labs and target applications are required. The working group would like to build a tri-lab debugger capability around the LLNL capability.

The Open Source Contract Maintenance effort provides funding to outside developers who maintain tools and tool infrastructures that are critical for code teams or serve as the basis for internal tools. This funding will be provided to those developers through support contracts administered mainly by LLNL, but each contract includes support for all three laboratories, and all three laboratories in close collaboration provide the technical guidance for the three contracts. This currently includes the tools, Open|SpeedShop, TAU, HPCToolkit, MUST/Vampir, and Valgrind.

The MPI integration/scaling effort is targeted to develop a set of capabilities focusing on supporting scale increases, assessing performance of both MPI and user applications, and providing optimal parameters to users for better MPI performance. The complex multi-socket, multi-core NUMA node architecture of systems such as the CTS-1 ~~TLCC2~~ mandates such an investigation. A close working relationship with the Open MPI community and other MPI developers is seen as a strategic need.

Required Capabilities

R1: Integration with system software environment to support increased scale

R2: Collaboration with Open Source tool providers and vendors to develop increased integration and leverage software to support increased scale

R3: Ability to support transition to new programming models and system architectures

R4: Support compilation, debugging, performance analysis, and monitoring

Gaps

G1: Ability to support programming model changes required for evolving architectures

G2: Ability to support increased computation scale

Five-Year Plan

FY14

- Evaluate tri-lab needs, with regard to performance analysis and debugging, with evolving MPI+X programming model(s), providing node-centric performance information, power utilization information, and information to help minimize data movement overheads (G1, G2)
- Provide enhancements and bug fixes to Open MPI/MVAPICH based on tri-lab need; assess MPI performance across many architectures; assess the impact of process and memory binding policies on application performance; and provide results to end users
- Deliver enhanced capabilities addressing user-identified needs for the TotalView debugger through the BIGCAT tri-lab collaboration
- Continue development and support efforts for debuggers, performance analysis tools, and MPI as programming models and architectures evolve (G1, G2)

FY15

- Evaluate tri-lab needs for Trinity-like architectures with a focus on the tools needed to improve application resiliency using "Burst Buffers;" investigate tools to analyze performance and debug applications modified for standard or defensive I/O with an objective to improve job throughputs for individual jobs as well as the system throughput with multiple jobs (G1, G2)
- Address programming environment issues required for successful deployment of the CTS-1 systems (G1, G2)
- ~~Continue to integrate programming model and architecture evolution into supported tools (G1, G2)~~

FY16

- Continue to integrate programming model and architecture evolution into supported tools (G1, G2)

FY17

- Continue to integrate programming model and architecture evolution into supported tools (G1, G2)

FY18

- Continue to integrate programming model and architecture evolution into supported tools (G1, G2)

~~*Resource Management Deployment and Reporting*~~

For FY14, this project is being integrated into the "System Software Deployment for Commodity Technology Systems" and the "High Performance Computing Environment Integration to Tri-Lab Systems" projects.

~~The Resource Management Deployment and Reporting project is focused on increasing the capability to better manage and schedule HPC resources, report on resource utilization, and build the tri-lab capability to support its growth and evolution toward larger scale.~~

~~The Resource Manager Capability Integration effort will increase the support level for SLURM (resource manager that is part of the CCE software stack) across the three laboratories. SLURM development is moving primarily to a commercial entity. While this is the model that ASC strives for with products originating at the labs, it is necessary to ensure the tri-labs needs are met as SLURM grows in use by other organizations. This project will integrate the support at each laboratory, including collection and coordination of requirement, and ensure that testing for the ASC workload and requirements is done.~~

~~The tri-lab Workload Characterization (WC) effort works closely with ASC HQ to support its reporting requirements for ASC HPC resources, using WC Tool. The tool provides the capability to report HPC resource requirements (as defined by each lab's ASC senior management) and utilization, with programmatic characterization of the work. WC Tool uses Resource Management SLURM/Moab capabilities to tie job requests and resulting platform usage data to their respective workload characterization. The tool includes a modular mechanism that can interface to multiple databases and work within computing environment differences at each of the three laboratories.~~

~~***Required Capabilities***~~

~~R1: Maintain optimal SLURM functionality across TLCC and tri-lab capability clusters~~

~~R2: Provide resource management capability to ensure highest computing resource utilization, ensure highest grade of service to users with minimal down times, collect and report system status and historical resource usage and utilization, respond effectively to system anomalies and user queries and complaints, and ensure continuation of capability as system software and hardware evolves and user needs become more complex and challenging~~

~~R3: Continue oversight and management of the Moab contract and the collaboration with Adaptive Computing~~

~~R4: Ensure WC Tool capabilities to effectively support ASC HQ reporting requirements~~

~~***Gaps***~~

~~G1: SLURM has grown to become increasingly complicated. Potential for problems to develop that exceeds our abilities to understand, diagnose, and solve.~~

~~G2: On most TLCC systems, SLURM is scheduled by Moab. Moab evolution may introduce disparities with SLURM, which will require attention adjustment.~~

~~G3: SLURM is open source but no longer a solely DOE-funded project. Tri-lab-specific requirements may develop that are not relevant or important to the worldwide SLURM community.~~

~~G4: Changes in tri-lab HPC computing environments may introduce reporting issues for WC Tool that will require attention.~~

~~Five-Year Plan~~

~~FY14~~

- ~~Continue SLURM development, installation, and support efforts through tri-lab collaboration (G1, G2, G3)~~
- ~~Continue WC Tool efforts to meet new and/or expanded ASC HQ reporting requirements; address issues in evolving tri-lab computing environments (G4)~~

~~FY15~~

- ~~Continue SLURM development, installation, and support efforts through tri-lab collaboration (G1, G2, G3)~~
- ~~Continue WC Tool efforts to meet new and/or expanded ASC HQ reporting requirements; address issues in evolving tri-lab computing environments (G4)~~

~~FY16~~

- ~~Continue SLURM development, installation, and support efforts through tri-lab collaboration (G1, G2, G3)~~
- ~~Continue WC Tool efforts to meet new and/or expanded ASC HQ reporting requirements; address issues in evolving tri-lab computing environments (G4)~~

~~FY17~~

- ~~Continue SLURM development, installation, and support efforts through tri-lab collaboration (G1, G2, G3)~~
- ~~Continue WC Tool efforts to meet new and/or expanded ASC HQ reporting requirements; address issues in evolving tri-lab computing environments (G4)~~

High Performance Computing Environment Integrations for Tri-Lab Systems

The HPC Environment Integration project targets the ability to work across sites with minimal transition and access restrictions. Differences in tri-lab security implementation and network restrictions as well resource access and authorization processes have been a hurdle. Efforts target ~~tri-lab security MOUs,~~ network access infrastructure, cross-realm authentication and resource management and environment standardization. Current efforts include:

- Inter-Site HPC targeted to establish a collaborative environment across sites: ~~based upon a tri-lab ISA and governance model:~~
  - Security policies and implementations that allow tri-lab access using identified protocols and technologies and to facilitate easier resource utilization; utilization, where feasible, of home site security apparatus, including authentication equipment (for example, single sign on)
  - User access to tri-lab resources within a specified period of time upon request and based on identified need
  - Support of multiple classes of users, such as architecture and modeling, application, customer service, system administrators with root access, and end users
  - Necessary network bandwidth and latency to cover current and projected requirements
- Establishing a cross-site authentication and resource approval through enhancements to the SARAPE system. This is a Web-based application that allows users within restricted domains to request selected CCE resources to which they are permitted access. It addresses the APIs required to help interface SARAPE with other tools required to manage accounts among the tri-labs. As part of the IHPC deployment, a service catalog will be deployed through which collaborators can view and request accounts and services available in the shared environment.
- The Shared Workspace effort is the infrastructure for promoting collaboration across the laboratories. It currently includes the Gforge server that is housed and managed at SNL.

Required Capabilities

R1: A tri-lab HPC environment that allows ease of cross-site usage; network and security policies; computing resources

R2: Systems and tools in place that collaborative project development, communication, and leveraging of resources

Gaps

G1: Differences in security policies and infrastructure across sites produce roadblocks

G2: Limited ability to manage users and resources from a cross-site perspective

G3: Limited ability for users to collaborate and move data between sites

G4: Changes in tri-lab HPC computing environments may introduce reporting issues for WC Tool that will require attention

Five-Year Plan

FY14

- Deploy expanded authentication capabilities to targeted services (G1, G2)
- ~~Based on results of FY13 investigation,~~ Test implementation of ~~shared file systems and~~ unclassified disaster recovery (G1, G3)
- Continue integration of collaborative tools as required (G3)
- Continue to integrate ~~improved authentication capabilities for~~ next SARAPE development phases; integrate improved authentication capabilities (G2)
- Continue WC Tool efforts to meet new and/or expanded ASC HQ reporting requirements; address issues in evolving tri-lab computing environments; deploy a new RAILS 3.x version of WC Tool

FY15

- ~~Implement into production shared file systems and unclassified disaster recovery (G1, G3)~~
- Develop revised working model across sites based on infrastructure; achieve site concurrence (G2, G3)
- Continue integration of collaborative tools as required, including SARAPE (G3)
- Continue WC Tool efforts to meet new and/or expanded ASC HQ reporting requirements; address issues in evolving tri-lab computing environments (G4)

FY16

- Deploy changes to site environments based on developed working model (G2, G3)
- Continue integration of collaborative tools as required, including SARAPE (G3)
- Continue WC Tool efforts to meet new and/or expanded ASC HQ reporting requirements; address issues in evolving tri-lab computing environments (G4)

FY17

- Continue changes to site environments based on developed working model (G2, G3)
- Continue integration of collaborative tools as required, including SARAPE (G3)
- Continue WC Tool efforts to meet new and/or expanded ASC HQ reporting requirements; address issues in evolving tri-lab computing environments (G4)

FY18

- Continue changes to site environments based on developed working model (G2, G3)
- Continue integration of collaborative tools as required, including SARAPE (G3)
- Continue WC Tool efforts to meet new and/or expanded ASC HQ reporting requirements; address issues in evolving tri-lab computing environments (G4)

Monitoring and Metrics Integrations for Tri-Lab Systems

The Monitoring and Metrics Integration project targets efficient and productive use of HPC systems as well as informed future planning through: 1) effective monitoring of all measurable or reportable conditions on compute platforms, both current and future, that can impact the performance of both applications and throughput on those platforms; and 2) appropriate transformation of monitored information into metrics and transport of those metrics to facilitate their use by system utilities, applications, resource managers, users, system administrators, and management. Integration of information from disparate data sources will enable greater system understanding and response to system conditions.

The project is targeting issues in current systems but with a focus on the increases in both scale and complexity that are expected over the next 5–10 years. With respect to software development, existing tools will be leveraged to the degree possible (for example, Splunk for appropriate data analysis tasks and Web protocols for information transport), and software will be written where gaps exist that cannot be filled by existing software (for example, distributed data services incorporates project-written software at the endpoints and existing transport frameworks).

The project currently focuses on three areas: 1) deployment of developed monitoring infrastructure (for example, Splunk) at all three sites, 2) development of a common scheme for organization, access, and aggregation of monitored information to improve troubleshooting and productivity, and 3) run-time use of platform resource utilization and state information by applications for increased performance.

Required Capabilities

R1: Efficient HPC troubleshooting through platform independent information aggregation and analysis tools

R2: Optimization of resource utilization through effective use of monitored information

R3: Portability of tools and infrastructure both cross-platform and cross-lab

Gaps

G1: Required information is typically unavailable as an aggregate due to distributed data sources and no standardized uniform data querying and transport mechanisms exist

G2: Scalable platform independent analysis tools that can utilize both text and numeric information as an aggregate do not exist

G3: Information/names change across platforms and applications

G4: Metrics for platform understanding are limited due to difficulty in obtaining a comprehensive view of platform state

G5: Shared repositories (cross-lab) for both software and documentation

G6: Metrics for understanding application/platform resource interaction are limited due to lack of low impact tools for collection and analysis as well as lack of exposure of related information from both hardware and system software

Five-Year Plan

FY14

- Continue development of Splunk tools as needed across the sites (LANL, LLNL, SNL) (G1, G5)
- Deploy LDMS for use as common data collection, transport, and storage tool on tri-lab HPC systems
- ~~Continued expansion of common information dictionary (G3, G5)~~
- ~~Explore what system and application related metrics (including power), both raw and derived, on what time-scales are germane to acceptable/increased application performance and system efficiency (G6)~~
- Begin investigation of how expected architectural changes and increased scale will impact the ability to effectively monitor and analyze information pertinent to application performance and system operation (G2, G6)
- Begin development of information aggregation tools that make use of the information sharing frameworks ~~and APIs developed in FY12 and FY13~~ (G1, G2, G4)

FY15

- Deploy into production common monitoring, analysis, and feedback tools on ASC platforms (G1, G3, G4, G5)
- Address mismatches in monitoring and analysis capabilities and new requirements due to evolution in application programming models, system software, and changes in scale and complexity of new systems (G2, G6)

FY16

- Continue to evolve monitoring, analysis, and feedback tools as systems and applications evolve (G1, G2, G3, G4, G5, G6)
- Test tools in new HPC environments (both production and experimental) to ensure readiness as new platforms arrive (G1, G2, G3, G4, G5, G6)

FY17

- Continue to evolve monitoring, analysis, and feedback tools as systems and applications evolve (G1, G2, G3, G4, G5, G6)
- Test tools in new HPC environments (both production and experimental) to ensure readiness as new platforms arrive (G1, G2, G3, G4, G5, G6)

FY18

- Continue to evolve monitoring, analysis, and feedback tools as systems and applications evolve (G1, G2, G3, G4, G5, G6)
- Test tools in new HPC environments (both production and experimental) to ensure readiness as new platforms arrive (G1, G2, G3, G4, G5, G6)

~~*File System Architecture and Integration*~~

For FY14, this project is being integrated into the "Next Generation Computing" project.

~~The File System Architecture and Integration project integrates tri-lab efforts regarding file system and data management deployment and in developing I/O middleware to speed up I/O for codes that run on tri-lab and other HPC resources. It leverages and supports other CSSE and FOUS efforts with a focus on production utilization.~~

~~This project initially focuses on leveraging on-going efforts, such as LANL's Parallel Log Structured File System (PLFS), and LLNL's Scalable/Check-Point Restart Code (SCR), to share knowledge, experience, and source code between the labs, to cooperate on development in areas of common needs including support for I/O burst buffers, to support standard APIs including POSIX and MPI I/O with minimal application changes, to work out issues involved in a wider production installation and support process, and to form the basis for working on a file systems architecture that these and other projects can build on. This supplements on-going efforts.~~

~~With current efforts outlined in the Fast Forward program and the wider interest in the community (both lab and commercial) of these technologies, this project will continue with integration efforts of PLFS/SCR and also take related prototype functionality from the Fast Forward program and test independently from a production environment focus. This provides a feedback process to other partners and keeps production needs in mind.~~

~~*Required Capabilities*~~

~~R1: Tri-lab integration support of file system infrastructure systems and strategic efforts~~

~~R2: Scalable I/O deployment based on PLFS/SCR efforts~~

~~*Gaps*~~

~~G1: Unknown location of burst buffer in future HPC architectures~~

~~G2: Characterizing the application problem space and identifying barriers that PLFS/SCR can address~~

~~G3: Identify methods to extract data from cluster, both node local and burst buffer, with minimal interference to application~~

~~G4: Identify "best" format to move and store data when using PLFS/SCR~~

~~G5: Existing application compatibility with Fast Forward architecture~~

~~*Five-Year Plan*~~

~~FY14~~

- ~~Investigate and potentially test support of PLFS index and data files reorganization during transfer (G2, G3)~~
- ~~Test Lustre Object Store/PLFS-like integration (G5)~~

- ~~Demonstrate burst buffer (G1, G2, G5)~~
- ~~Investigate current applications using end-to-end epoch recovery, including SCR (G5)~~
- ~~Test Fast Forward APIs for sufficiently large and varied datasets (G4)~~

~~FY15~~

- ~~Examine API efforts needed to support future distributed storage (G1, G2, G5)~~
- ~~Work with the CCE Resource Manager project to assess and test resource manager integration of file system services (based on Fast Forward project plans) (G2, G5)~~

~~FY16~~

- ~~Assess all related file-system- and I/O-related roadmaps (G1, G5)~~
- ~~Continue development and testing of I/O and integrated infrastructure (G2, G3)~~

~~FY17~~

- ~~Continue development and testing of I/O and integrated infrastructure (G2, G3)~~